REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	23 Oct 95	

4. TITLE AND SUBTITLE
Response Surface Analysis of Two-Stage Stochastic Linear Programming with Recourse

5. FUNDING NUMBERS

6. AUTHOR(S)
Thomas Glenn Bailey

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
AFIT Student Attending:
University of Texas at Austin

8. PERFORMING ORGANIZATION REPORT NUMBER
96-011D

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DEPARTMENT OF THE AIR FORCE
AFIT/CI
2950 P STREET, BLDG 125
WRIGHT-PATTERSON AFB OH 45433-7765

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release IAW AFR 190-1
Distribution Unlimited
BRIAN D. GAUTHIER, MSgt, USAF
Chief Administration

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

19960603 015

14. SUBJECT TERMS

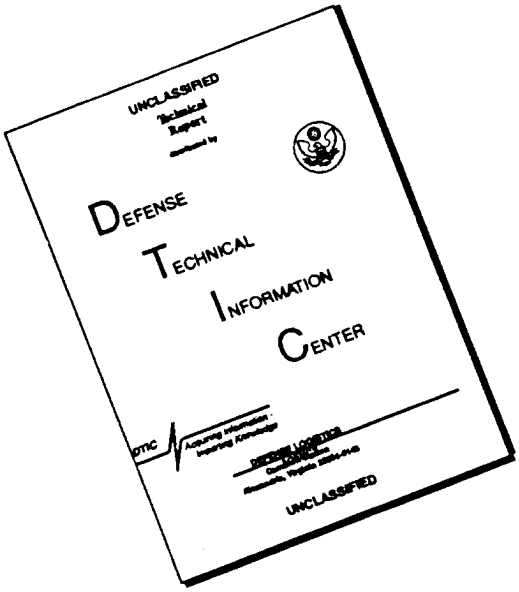
15. NUMBER OF PAGES
241

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

DTIC QUALITY INSPECTED 1

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	**PR**	- Project
G	- Grant	**TA**	- Task
PE	- Program Element	**WU**	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- **DOD** - See DoDD 5230.24, "Distribution Statements on Technical Documents."
- **DOE** - See authorities.
- **NASA** - See Handbook NHB 2200.2.
- **NTIS** - Leave blank.

Block 12b. Distribution Code.

- **DOD** - Leave blank.
- **DOE** - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- **NASA** - Leave blank.
- **NTIS** - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

★U.S.GPO:1990-0-273-271

Copyright

by

Thomas Glenn Bailey

1995

Response Surface Analysis of Two-Stage Stochastic Linear Programming with Recourse

by

Thomas Glenn Bailey, B.S., M.A., M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 1995

Response Surface Analysis of Two-Stage Stochastic Linear Programming with Recourse

Approved by Dissertation Committee:

Paul A. Jensen, Supervisor

To Beverly, Becky, and Kenny

Acknowledgments

I wish to express my deepest gratitude to Dr. Paul Jensen for his supervision, patience, dedication, and encouragement during all phases of my doctoral program at the University of Texas. This dissertation would not have happened without his early and continuing interest and assistance, for which I'm most grateful. I also extend my sincere appreciation to Drs. J. Wesley Barnes, Jonathan Bard, Leon Lasdon, and Doug Morrice for their suggestions and guidance as both instructors and committee members; and, a special thanks to Dr. David Morton for his extra efforts in supporting this research.

I would also like to acknowledge the help of numerous faculty, staff, and students for the many ways — direct or otherwise — they contributed to helping me complete this research, especially Lt. Col. Bee Carlton, Dr. Valerie Tardif, Sue Ponder, and Astrid Leuba. I also wish to thank the staff of the Air Force Military Personnel Center and the Air Force Institute of Technology for giving me this opportunity.

Most importantly, I thank my wife Beverly, and children Becky and Kenny, for their love, patience, and sacrifice throughout this endeavor. This accomplishment is theirs.

T.G.B.

The University of Texas at Austin

Austin, Texas

October 23, 1995

Response Surface Analysis of Two-Stage Stochastic Linear Programming with Recourse

Publication No._____

Thomas Glenn Bailey, Ph.D.

The University of Texas at Austin, 1995

Supervisor: Paul A. Jensen

This research investigates a special class of stochastic linear programs known as two-stage stochastic linear programming with relatively complete and fixed recourse. These models characterize a two-phase process where the first-stage decision (itself subject to a separate set of first-stage linear constraints) allocates a set of resources to the second-stage linear program prior to the realization of random variables affecting second-stage resource availability. Since the second-stage decision deterministically follows both first-stage allocation and random variable realization, the first-stage variables constitute the only true decision. The expected cost of the two-stage recourse problem is also a piecewise convex function of the first-stage decision variables, thus allowing a global optimal solution that minimizes the total expected cost.

This dissertation extends previous efforts on finding this optimal solution by introducing new optimization algorithms that take advantage of recurring optimal bases in the recourse problem. Additionally, the convergence and efficiency of the Geometric Simplex, Projected Gradient, and Parallel Tangents optimal search techniques are explored on a set of five capacity expansion problems from the literature involving power generation and vehicle basing.

This research also applies the techniques of experimental design and response surface methodology to approximate the expected cost as a quadratic polynomial function of the first-stage variables. Using canonical analysis, this dissertation provides directions of minimum and maximum response sensitivity to deviations in the optimal values of the first-stage variables that were previously unavailable. This research demonstrates the practicality of such analysis on problems with up to 63 first-stage decision variables and over $1.099 \cdot 10^{12}$ recourse scenarios.

This dissertation also introduces a new type of variance reduction — Latin Hypercube sampling — to this class of problems by showing it guarantees a reduction in the variance of the estimators of the expected cost, and empirically confirming its consistently large variance reduction through comparisons with random sampling and control variate results. Finally, this research introduces tolerance limits as a non-parametric-based technique for characterizing the underlying distribution of the recourse problem.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Problem Overview

Most practical problems encountered by operations research practitioners intrinsically possess some level of uncertainty within their own parameters. While stochastic-based modeling techniques inherently accommodate such randomness, they typically cannot *prescribe* an optimal course of action. By contrast, well-defined optimization methods can find an optimal solution, but only under the simplifying assumption of deterministic parameters (usually expected values). Unfortunately, neither modeling paradigm can — by itself — give a decision maker the optimal course of action that hedges against extreme realizations of the random parameters.

Long recognizing this deficiency, researchers from both camps have attempted to bridge the two by incorporating elements of one within the operational framework of the other. For stochastic-oriented methods, this migration has occurred primarily under numerical approximation methods, typically by applying optimization techniques such as gradient search and response surface analysis to a simulation model of the problem. For mathematical programming-based algorithms, incorporating stochastic elements has led to algorithms involving parameter approximations, scenario representation through large-scale models, iterative sampling, and parallel processing. In both cases, the computational requirements can be formidable.

It is in this spirit that this dissertation offers a new approach to analyzing the well-defined problem of stochastic programming with recourse. Strictly speaking, as a Monte Carlo simulation this research falls in the stochastic-based category of modeling techniques. However, the underlying principles that motivate the simulation come from linear programming theory; mathematical programming literature dominates previous research in recourse optimization problems; and, most recourse problems available for comparison purposes have been solved using LP-based methods. Consequently, this chapter and the following literature review will initially proceed from the mathematical programming perspective.

1.2 PROBLEM DESCRIPTION

The literature offers a diverse class of problems under the category of stochastic linear programming (SLP) based on several properties.

1. *Random Parameter Location*. This characteristic of SLP concerns which parameters of the model (objective, constraint matrices, or right-side vectors) are random variables.

2. *Probabilistic Representation*. This aspect of SLP concerns the modeling issue of the random variables; i.e., their portrayal either through probability distribution functions or by representative scenarios. Furthermore, this issue addresses distributional assumptions of type and

parameter, or in the case of scenario modeling the sampling of a wide variety of possible outcomes.

3. *Structural Assumptions*. This feature of SLP relates to the number of time periods associated with the realization of the random variables; type and availability of recourse available to the decision-maker after such realizations; chance constraints; and, assumptions on the separability of decision variables or scenarios.

4. *Constraint Flexibility*. This form of SLP allows for the possibility of constraint violation, either through predetermined levels of probability or through explicit error vectors to account for infeasibility.

One category of SLP that was recognized and reviewed early on (Dantzig 1955), and has continually reappeared in the literature, is the two-stage problem with fixed and complete recourse:

$$\text{MIN } Z(\mathbf{x}) = \mathbf{cx} + E[h(\mathbf{x},\omega,\mathbf{T})], \text{ s.t. } \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}, \tag{1.1a}$$

where

$$h(\mathbf{x},\omega,\mathbf{T}) = \text{MIN } \mathbf{dy}, \text{ s.t. } \mathbf{Wy} = \omega - \mathbf{Tx}, \mathbf{y} \geq \mathbf{0}, \tag{1.1b}$$

and ω and $\mathbf{T}$ contain one or more elements that are random variables with prescribed distributions. This form of stochastic programming represents many practical problems where the uncertainty lies in resource demand or consumption; and, in recent years has been the research focus regarding advanced

decomposition algorithms involving Monte Carlo sampling, statistical measures of ω and $\mathbf{T}$, and convergence. However, the literature's approach to solving (1.1) has exclusively focused on *finding* MIN $Z(\mathbf{x})$, while little has been reported regarding ***solution sensitivity*** or the ***underlying distribution*** with respect to $h(\mathbf{x},\omega,\mathbf{T})$. As a practical matter, though, such analysis provides the decision-maker with the insight to choose a solution that incorporates subjective assessments that are not explicitly modeled. Furthermore, the distributional properties of (1.1b) characterize the behavior of the recourse function in a way that gives the decision-maker insight beyond the sole criteria of expected value.

Justification for this line of inquiry begins with the three primary characteristics known to be true for (1.1).

1. *Decision Variables.* The structure of (1.1) dictates that $\mathbf{x}$ represents the only true decision vector; once made, the recourse problem (1.1b) becomes a deterministic function of $\mathbf{x}$ and the realization of the random variables in $\mathbf{T}$ and ω. In a simulation context, this result places $\mathbf{x}$ as the independent variable and $Z(\mathbf{x})$ as the dependent response. Furthermore, preliminary and past research provides empirical evidence that the region around the optimal solution $\mathbf{x}^*$ is often 'flat'; i.e., for some small epsilon value $\varepsilon > 0$ there exists a considerable range of decision variables $\mathbf{x}'$ such that $Z(\mathbf{x}') = Z(\mathbf{x}^*) + \varepsilon$ (including multiple optimal solutions; i.e. $\varepsilon = 0$).

2. *Convexity.* The literature shows that $Z(\mathbf{x})$ is a piecewise linear convex function of $\mathbf{x}$, a property that holds several important ramifications from a

simulation perspective. First, such convexity suggests approximating $Z(\mathbf{x})$ with a quadratic polynomial formulation of $\mathbf{x}$. Second, a quadratic assumption greatly reduces the size of the experimental design used to collect the data. Finally, knowing the general shape of the response allows for an independent verification of the validity of the polynomial approximation using its eigenvalues — a capability generally not available in response surface methodology due to the unknown underlying functional form.

3. *Distributional Form.* Both the literature and preliminary research show that the distribution of $h(\mathbf{x},\omega,\mathbf{T})$ varies with $\mathbf{x}$. Consequently, there may be cases where $\mathbf{x}^*$ *is not be the best answer* if the distribution of $h(\mathbf{x}^*,\omega,\mathbf{T})$ is less favorable in certain aspects (e.g., form, variance, range) relative to other parts of the feasible region.

Therefore, this dissertation proposes expanding the analysis of (1.1) beyond just finding the optimal solution by (1) deriving low-order polynomial approximations to $Z(\mathbf{x})$ in the region of optimality, and (2) comparing important characteristics of the distributions of $h(\mathbf{x},\omega,\mathbf{T})$ for those values of $\mathbf{x}$ in the region of optimality.

This research will focus on a special and important class of (1.1) — *the capacity expansion problem* — and examine its results on a collection of problems provided by Morton (1994c) and downloaded through the INTERNET (http://www.engin.umich.edu/~dholmes). Table 1.1 gives a description of the size of the problem sets denoted as APL1P, PGP2, CEP1, 4TERM, and 20TERM.

TABLE 1.1
PROBLEM SET DESCRIPTIONS

Problem Set	# of $\mathbf{x}$ Variables	# of Random ω/T	# of Scenarios	Rows/Cols in $\mathbf{A}^{\dagger}$	Rows/Cols in $\mathbf{W}$
APL1P	2	5	1280	2 / 2	5 / 9
PGP2	4	3	576	2 / 4	7 / 16
CEP1	4	3	216	9 / 8	7 / 15
4TERM	15	8	256	3 / 15	28 / 146
20TERM	63	40	$1.095 \cdot 10^{12}$	3 / 63	124 / 764

† - Does not include upper bounds on $\mathbf{x}$.

With the exception of APL1P, Chapter 5 provides the detailed description and analysis results of these problems. Chapters 3 and 4 use APL1P as a demonstration problem due to its small size.

1.3 DISSERTATION OUTLINE AND CONTRIBUTIONS

Current approaches that use sampling techniques to solve (1.1) share the same basic philosophy of incorporating such methods within a linear programming framework. *However, to find the response surface approximations of (1.1) this dissertation proposes a different modeling framework by using linear programming results within a Monte Carlo simulation environment.* Although such an approach provides a far richer analysis, it comes at the expense of increased computational burdens. Consequently, this shift in modeling paradigm requires using a variety of techniques to efficiently solve recourse problems, and can be classified in two categories: *Optimization* and *Statistical Analysis*.

Figure 1.1 breaks these two categories into six principal topics and shows how they fit in the overall organization of this dissertation. This research provides its major contributions within these topics in the following manner.

1. *Search Techniques.* This dissertation compares three different optimal search routines — *Geometric Simplex*, *Projected Gradient*, and *Parallel Tangents* — for accuracy and convergence. The simplex and parallel tangent approach have not been tried before in the recourse context, while the projected gradient method has been discussed in the general case of stochastic quasigradient methods.

2. *Optimization Algorithms.* Extending previous research in the areas of basis classification, this thesis completely enumerates the optimal bases for either the entire feasible region or prescribed subset. This new technique offers a faster way to calculate estimates of $Z(\mathbf{x})$, and for smaller problems offers an order of magnitude increase in efficiency.

3. *Variance Reduction.* The literature reports limited use of variance reduction techniques (primarily importance sampling) to improve the accuracy, or ease the computational burden, of estimating $Z(\mathbf{x})$. This research investigates two techniques — *Control Variates* and *Latin Hypercube*.

4. *Experimental Design.* This topic represents the first known attempt to analyze recourse problems using experimental design methods from the simulation and engineering literature.

	Search Technique	*Optimization Algorithm*	*Variance Reduction*	*Experimental Design*	*Response Surface Analysis*	*Distribution Analysis*
	CHAPTER 2 LITERATURE REVIEW					
Literature Review	• Background • Decomposition • Quasigradient	• Theory • Assumptions, Observations, and 'Folklore'	• Importance Sampling	*Not Applicable*	*Not Applicable*	*Not Applicable*
	CHAPTER 3 METHODOLOGY - OPTIMIZATION		CHAPTER 4 METHODOLOGY - STATISTICAL ANALYSIS			
Methodology	• Search Method •• Simplex •• Gradient •• PARTAN	• Algorithms •• OSL •• OBS-CMPL. •• OBS-RESET •• ODV	• Control Variates • Latin Hypercube	• Screening • Composite • Fractional	• Canonical Analysis • Ridge Analysis	• Distributional Characteristics • Tolerance Limits
	CHAPTER 5 PROBLEM SET ANALYSIS					
Results	• Speed • Accuracy • Convergence	• Comparison to OSL	• Amt. of Var. Reduction • Improved Search/Design	• Quadratic Fit • Positive Definite	• Min/Max Sensitivity • Define Optimal Region	• Statistical Analysis of Optimal Region

Figure 1.1. Dissertation Overview

5. *Response Surface Analysis.* This topic employs *Canonical Analysis* as the principal method for quantifying the region of optimality, and directions of minimum and maximum sensitivity.

6. *Distributional Analysis.* Using non-parametric methods, this research introduces this concept to recourse problem analysis by using *Tolerance Limits* to consider high-cost scenarios.

As shown in Figure 1.1, Chapter 2 reviews the literature on stochastic linear programming and simulation optimization, with a special emphasis on solution techniques for two-stage recourse problems. Chapter 3 presents the methodology and underlying theory employed by the proposed solution algorithms and search techniques. Chapter 4 covers the statistical analysis topics and their application in the context of (1.1). Chapter 5 reviews the capacity expansion problem sets found in the literature and on the INTERNET, and provides a response surface and distributional analysis for each problem using the proposed algorithm. Chapter 6 summarizes the results and contributions of this dissertation, and concludes with proposals for future research.

Chapter 2

Literature Review

2.1 INTRODUCTION

The literature on stochastic optimization can be roughly categorized into two broad arenas — stochastic programming and simulation. This dissertation proposes solving the two-stage recourse problem traditionally considered within the domain of mathematical programming under uncertainty by synthesizing results and techniques developed in both camps; accordingly, this chapter reviews previous applicable investigations in these areas in the following manner. Section 2.2 surveys the different modeling variations and assumptions found in the literature under the general heading of stochastic linear programming, with particular emphasis on the category of interest — two-stage programming with recourse. Section 2.3 reviews proposed solution techniques for two-stage recourse problems, while Section 2.4 closes the chapter with a summary of the simulation optimization literature.

2.2 STOCHASTIC LINEAR PROGRAMMING

2.2.1 General Description

Stochastic linear programming was first explored by Dantzig (1955) as mathematical programming under uncertainty, and independently investigated by Beale (1955). Since then, the term *stochastic programming* has come to encompass a very broad range of mathematical programming problems where

uncertainty exists in one or more of the parameters. Wets distinguishes between decision making under uncertainty, where little is known quantitatively about the uncertainty, and decision making under risk (or stochastic programming) as problems where "...the decision maker is given a description of these unknown parameters in terms of a well-defined probability law (Wets 1974)." For the purposes of this research, the term *stochastic linear programming* (SLP) follows Kall's (1976) definition as being "...concerned with problems arising when some or all *coefficients* of a linear program are *stochastic variables* with *known (joint) probability distribution.*"

A generalized version of SLP given by Ermoliev and Wets shows the difficulty of solving stochastic programs with recourse:

$$\begin{aligned} &\text{find} && x \in X \subset R^N \\ &\text{such that} && F_i(x) = E\{f_i(x,\omega)\} \\ &&& \quad = \int f_i(x,\omega)P(d\omega) \le 0, \ i = 1, \ldots, m \\ z &= F_0(x) = && E\{f_0(x,\omega)\} \\ &&& \quad = \int f_0(x,\omega)P(d\omega) \text{ is minimized} \end{aligned} \tag{2.1}$$

where (Ω,A,P) define the probability space, and for every x in X, f_i, f_0, and all expectations are defined. The simplest approach for solving (2.1) — *scenario analysis* — finds a optimal solution $x^*(\omega')$ for a given scenario ω'; however, where there exists ω^k $(k = 1, \ldots, s)$ scenarios to consider, a convex combination of $x^*(\omega^k)$ $(k = 1, \ldots, s)$ may prove to be infeasible (Ermoliev and Wets 1988). *Scenario Optimization*, as proposed by Dembo (1989), extends this approach to linear forms of (2.1) by compensating for such possible infeasibilities by solving

(2.1) for each scenario k, then optimizing a *tracking* or *coordination* model that minimizes some function of the differences in optimality and feasibility; e.g.,

$$\begin{aligned} \text{MIN} \quad & \sum_{k=1}^{s} p_k \|\mathbf{c}_k\mathbf{x} - v_k\|^2 + \sum_{k=1}^{s} p_k \|\mathbf{A}_k\mathbf{x} - \mathbf{b}_k\|^2 \\ \text{s.t.} \quad & \mathbf{A}_d\mathbf{x} = \mathbf{b}_d, \qquad \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{2.2}$$

where $\mathbf{A}_k\mathbf{x} = \mathbf{b}_k$ characterizes the constraint set for scenario k, $\mathbf{A}_d\mathbf{x} = \mathbf{b}_d$ represents deterministic constraints for all scenarios, and v_k is the optimal objective function value for scenario k (Dembo 1989, 1991). However, many problems require *non-anticipitivity*, or *here-and-now*, decisions, i.e., choices that must be made once-and-for-all prior to the realization of ω. In these cases scenario analysis or scenario optimization models are inappropriate (Ermoliev and Wets 1988, Morton 1994b).

Ermoliev and Wets also contrast the here-and-now environment (which requires an *anticipative* optimization solution) to the situation where the decision maker observes ω *prior* to deciding x. They term this type of model *adaptive optimization*, and in particular introduce the *distribution problem* as one of finding the distribution function of the optimal value of adaptive models. They also introduce the two-stage recourse model "...as an attempt to incorporate the fundamental mechanisms of anticipation and adaptation within a single mathematical model. In other words, this model reflects a trade-off between long-term anticipatory strategies and the associated short-term adaptive adjustments (Ermoliev and Wets 1988)".

Because of their dual anticipative/adaptive nature, recourse formulations present a robust mechanism for analyzing stochastic programming problems. Furthermore, the inherent uncertainty of these problems implies a need for estimating the distribution function of the optimal solution, the confidence interval of the point estimates of $x^*(\omega)$, and the objective function value. Although such estimates are extremely difficult to find (Ermoliev and Wets 1988), this research proposes a technique for such an undertaking for a special case of recourse problems. Specifically, the research focuses on solving this area of SLP known as the *distribution problem* with respect to stochastic linear programming problems with *relatively complete recourse*. The problem set contains a specific and important class of recourse problems that model *capacity expansion*. This section reviews the appropriate literature for this class of stochastic linear problems and provides the context in which this research is conducted. In addition to Ermoliev and Wets (1988), general introductions and overviews of the field can be found in Birge and Mulvey (1994), Dempster (1980), Hansotia (1980), Kall (1976), Kall and Wallace (1994), and Stancu-Minasian and Wets (1976).

2.2.2 Stochastic Linear Programming with Recourse

Dantzig (1955) first proposed what the literature now calls stochastic linear programming with recourse. In the general recourse case, a first-stage decision must be made without knowledge of the values of a subset of the problem parameters. A second-stage (recourse) decision follows after making the

first-stage decision and realizing the random variable(s). Mathematically, the two-stage recourse version of (2.1) can be expressed as

$$\begin{aligned} &\text{MIN } Z(\mathbf{x}) = \mathbf{cx} + E[h(\mathbf{x},\omega,\mathbf{T})] \\ &\text{s.t. } \mathbf{x} \in \boldsymbol{X}, \boldsymbol{X} = \{\mathbf{x} : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\} \end{aligned} \tag{2.3a}$$

where

$$\begin{aligned} h(\mathbf{x},\omega,\mathbf{T}) &= \text{MIN } \mathbf{dy} \\ \text{s.t. } \mathbf{Tx} + \mathbf{Wy} &= \omega, \mathbf{y} \geq \mathbf{0}; \end{aligned} \tag{2.3b}$$

x and **y** are the first- and second-stage decision variable vectors, respectively; **c** and **d** the respective cost vectors for **x** and **y**; **A** and **b** the matrix of technological coefficients and right-side resource vector, respectively, for the first-stage problem; and **W** and ω the *recourse* matrix and *random* right-side resource vector, respectively, for the second-stage problem. The matrix **T** (which in certain formulations can contain random components) represents the amount of resource consumed in the second-stage based upon the first stage decision **x**, and E is the expectation operator. The objective in (2.3a) is to find $\mathbf{x}^* \in \boldsymbol{X}$ such that MIN $Z(\mathbf{x}) = Z(\mathbf{x}^*)$.

There exists numerous variations in the literature on the structure of the recourse problem in (2.3) due to differences in the number and type of stochastic parameters, distributional assumptions, type of recourse available, and the presence of integer decision variables. Walkup and Wets (1967), and Wets (1974) provide one set of classifications based on recourse type and availability, and on the location of the random parameters.

1. *Simple Recourse.* This type of recourse problem differs from (2.3) in that while the decision vector **x** is still decided prior to the realization of the random variables, the formulation assumes a simpler recourse **W** = [**I**, -**I**].

2. *Fixed Recourse.* This term refers to the formulation given in (2.3), where the resource vector ω and matrix **T** may be random, but **W** must be *fixed.*

3. *Complete Recourse.* This condition implies that the second-stage problem has a feasible solution for *any* right-side value. A relaxation of this condition — *Relatively Complete Recourse* — indicates that a feasible second-stage solution exists for any feasible first-stage solution **x**.

This dissertation restricts its research to the class of problems of the form (2.3) that possess relatively complete recourse.

2.2.3 Chance-Constrained Stochastic Programming

Although not directly addressed by this research, several important variations of SLP should be briefly noted. One such class of models is *chance-constrained programming* first proposed by Charnes and Cooper (1959), which views the problem as finding a solution that does not exceed a given probability of violating one or more constraints. Mathematically, the chance-constrained version of (2.1) can be expressed as

$$\begin{array}{lll} \text{MIN} & \mathbf{cx} & \\ \text{s.t.} & \text{PROB}\{\mathbf{Ax} \geq \mathbf{b}\} \geq \alpha & \\ \text{or} & \text{PROB}\{\mathbf{A}^i\mathbf{x} \geq \mathbf{b}^i\} \geq \alpha^i \qquad \mathbf{x} \geq \mathbf{0} \qquad i = 1, \ldots, m & (2.4) \end{array}$$

where α or α^i are the confidence coefficients for the entire constraint set or the i^{th} constraint, respectively, and **A**, and **b** can be all, or in part, stochastic. Kall (1976) shows that chance-constraint and two-stage recourse problems are not always equivalent due to the fact that, unlike recourse programs, chance-constrained programs are not, in general, convex programming problems. Furthermore, he shows that even if the convexity condition exists for (2.4) the computational requirements remain formidable. Further information on chance-constraint programming can be found in Gartska (1980), Kirby (1970), Prekopa (1970), and Vajda (1980).

2.2.4 Robust Optimization

Mulvey, Vanderbei, and Zenios (1991) suggest another form of stochastic programming that includes an explicit set of error vectors in the recourse formulation; i.e.,

$$\begin{aligned} h(\mathbf{x},\mathbf{S},\mathbf{T}) = \text{MIN} \quad & \sigma(\mathbf{y}) + \rho^+(\mathbf{z}^+) + \rho^-(\mathbf{z}^-) \\ \text{s.t.} \quad & \mathbf{W}\mathbf{y}^i + \mathbf{z}^+ - \mathbf{z}^- = \mathbf{S}^i - \mathbf{T}\mathbf{x} \qquad \text{for i} = 1, \ldots, s \\ & \mathbf{y} \geq \mathbf{0}. \end{aligned} \tag{2.5}$$

The error vectors $\mathbf{z}^+$ and $\mathbf{z}^-$ in (2.5) provide recourse in addition to **W**; however, the penalty functions $\rho^+(\bullet)$ and $\rho^-(\bullet)$ can be adjusted to insure that $\mathbf{z}^+$ and $\mathbf{z}^-$ enter the basis only if there does not exist any **y** such that $\mathbf{W}\mathbf{y} = \mathbf{S} - \mathbf{T}\mathbf{x}$ and $\mathbf{y} \geq \mathbf{0}$. In effect, this extension — called *robust optimization* — guarantees the condition of relatively complete recourse. Their other advancement includes formulating higher moments than the expected value in the objective function for discrete

versions through the function $\sigma(\bullet)$. By weighting the variance term in (2.5) they construct an efficient frontier that describes a robust tradeoff between risk and reward (Mulvey Vanderbei, and Zenios 1991).

2.3 Solution Techniques for SLP with Recourse

2.3.1 Approximations and Bounds

Much of the research since Dantzig (1955) and Beale (1955) independently proposed the recourse problem focuses on solution techniques to (2.3). Morton provides a concise categorization of these methods into three areas: *Exact Solution, Approximation and Bounding, and Sampling* (Morton 1994a). Morton metaphorically compares these approaches to how we solve integration problems: (1) first, we would typically try to solve an integration problem analytically to get the *exact solution*; (2) next, for a more difficult integral we would attempt a numerical *approximation* such as Simpson's Rule; and, (3) finally by Monte Carlo *sampling* methods (Morton 1994b). Although exact methods are preferred, the literature indicates that such an approach is rarely practical. Consequently, approximation and bounding techniques become necessary.

Unfortunately, for most situations developing an approximation is necessary — and difficult. In their introduction to approximation techniques for stochastic programming Kall, Ruszczynski, and Frauendorfer (1988) note that the difficulty of solving the integral forms in (2.1) directly has lead to approximation methods for the vast majority of cases where certainty equivalents or small finite distributions do not exist. A review of the literature shows that approximation and bounding algorithms dominate the research community's efforts to solve

stochastic linear programming, with particular focus on efficient implementations of decomposition algorithms, parallel optimization, sampling procedures, stopping rules, and bounding.

Regarding approximation techniques, Kall, Ruszczynski, and Frauendorfer note for the general case of (2.1) that all such methods share three fundamental issues: (1) first, any approximation approach must correctly substitute ω with a discrete representation; (2) next, it must develop some measure of its accuracy; and, (3) it should provide a way to surpass that accuracy by finding a better approximation of the original random vector ω. They point out that accomplishing the first task requires sampling the probability space to find an accurate discretized version of ω, a task made difficult for several reasons. First, there exists the fundamental problem of not knowing beforehand how much is enough; e.g., how detailed a partition of the probability space is necessary. Second, the degree of non-linearity influences the degree of partitioning necessary for accurate approximation. Finally, the properties of $f_0(x)$ vary with $\boldsymbol{x}$, hence, the sampling itself depends on $\boldsymbol{x}$ (Kall, Ruszczynski, and Frauendorfer 1988).

Much of the recent literature on bounding algorithms for the recourse problem devotes its efforts to improving the error estimate associated with the upper bound. This effort stems from the fact that while calculating the lower bound under Jensen's inequality generally requires only a small number of evaluations for convergence, the upper bound calculations require evaluating $h(\mathbf{x},\omega,\mathbf{T})$ at every extreme point of the probability space, which for an m-random components in ω and $\mathbf{T}$ requires solving $h(\mathbf{x},\omega,\mathbf{T})$ 2^m times (Birge and Wets 1989). Additional problems stem from assumptions of independence; however,

both the empirical evidence and theoretical results imply that Jensen's inequality provides a better estimate of the optimal value of the recourse problem (Gassmann and Ziemba 1986). Consequently, the literature offer several extensions of these basic approximations.

Birge (1985) proposes aggregating rows and columns of the original recourse problem to reduce the computational complexity. Frauendorfer (1988) supplies a straightforward extension of the Edmundson-Madansky upper bound inequality to the case of dependent distributions among the elements of ω. Birge and Wets (1986) provide an extensive catalogue of approximation methods based upon linearization of an objective function (called *original*, *subgradients*, *rays*, and *pairs*) and one of several techniques for obtaining discrete probability measures (called *conditional expectations*, *extreme points*, *extremal probability measures*, and *majorizing probability measures*), together with guidelines for applying them to fixed recourse problems in conjunction with optimization solution methods. Gassmann and Ziemba (1986) suggest using linear programming on a partition of the probability space that provides a tighter upper bound, a technique extended by Edirisinghe and Ziemba (1992) to include variable and constraint aggregation, and multi-stage recourse applications. Birge and Wallace (1988), using the ray approximation procedure from Birge and Wets (1986), develop upper-bound approximations whose computational requirements are polynomial in the dimensionality of ω (also see Birge and Wets 1989). Finally, Birge and Dulá (1991) propose an extension of this sublinear approximation to include non-linear recourse problems with first- and second-moment information on the elements of ω.

A related area of interest regarding the upper and lower bounds of the recourse problem deals with the value and state of information regarding the uncertain parameters. Specifically, the literature addresses two fundamental questions in this area: (1) the value of additional information regarding the uncertainty in ω; and, (2) the degree of error between a deterministic approximation and its more accurate stochastic counterpart. First addressed by Madansky (1960) and Avriel and Williams (1970), Birge (1982) offers a summary of their work on *expected value of perfect information* (EVPI) and presents the *value of the stochastic solution* (VSS). Rewriting (2.3), Birge reviews previous results showing that for the problem

$$\begin{aligned} \phi(\mathbf{x},\omega,\mathbf{T}) &= \mathbf{cx} + \text{MIN}[\mathbf{dy} \mid \mathbf{Tx} + \mathbf{Wy} = \omega, \mathbf{y} \geq \mathbf{0}] \\ &\text{s. t. } \mathbf{x} \in \boldsymbol{X}, \boldsymbol{X} = \{\mathbf{x} : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}, \end{aligned} \tag{2.6}$$

where the expected value for the wait-and-see solution (WS) is

$$\text{WS} = \text{E}[\text{MIN}\ \phi(\mathbf{x},\omega,\mathbf{T})], \mathbf{x} \in \boldsymbol{X}, \tag{2.7}$$

the expected value for the recourse problem (RP) is

$$\text{RP} = \text{MIN}\ [\text{E}\ \phi(\mathbf{x},\omega,\mathbf{T})], \mathbf{x} \in \boldsymbol{X}. \tag{2.8}$$

The expected value approximation (EV) is defined as

$$\text{EV} = \text{MIN}\ \phi(\mathbf{x},\text{E}[\omega],\text{E}[\mathbf{T}]), \mathbf{x} \in \boldsymbol{X}, \tag{2.9}$$

and where $\bar{\mathbf{x}}$ is the optimal solution to (2.9), the expected result of using $\bar{\mathbf{x}}$ is

$$EEV = E[\phi(\bar{\mathbf{x}},\omega,\mathbf{T})], \mathbf{x} \in \boldsymbol{X}. \tag{2.10}$$

Consequently, the following bounds

$$EEV \geq RP \geq WS \geq EV \tag{2.11}$$

hold due to the convexity of $\phi(\mathbf{x},\omega,\mathbf{T})$ and Jensen's inequality. From (2.11) Birge shows that

$$\begin{aligned} EVPI &= RP - WS \\ EEV - EV &\geq EVPI \geq 0 \end{aligned} \tag{2.12}$$

and from Avriel and Williams (1970) repeats the suggestion that (2.12) provides a bound on the value of more information regarding ω and $\mathbf{T}$. Birge then suggests another measure — the value of the stochastic solution (VSS) — defined as

$$\begin{aligned} VSS &= EEV - RP \\ EEV - EV &\geq VSS \geq 0 \end{aligned} \tag{2.13}$$

to establish the worth and value bounds for solving more complicated recourse models (Birge 1982). Additional research on information costs includes ω and $\mathbf{T}$ with discrete distributions (Baron 1971), lower bounds for EVPI (Morris and Thompson 1980), and bounds for linear and concave preference functions (Hausch and Ziemba 1983).

2.3.2 Exact Solution Methods

According to Morton's classification, exact solution methods "...include simplex-based algorithms that exploit special structure of bases...decomposition or L-shaped schemes...interior point methods...and the Progressive Hedging algorithm...(Morton 94a)." To understand how and why these solution methods can be applied to (2.3), its general characteristics (such as convexity) need to be determined. Wets (1966a) accomplished this characterization by showing that (2.3a) is convex and continuous for that subset of solutions $\mathbf{x} \in \boldsymbol{X}$ when (2.3a) is feasible for all realizations of ω and $\mathbf{T}$. Furthermore, he shows that the expected value of $\pi(\omega - \mathbf{Tx})$ can be used to construct a supporting hyperplane to (2.3a) (Wets 1966a). These important results theoretically clear the way for solving (2.3a) with variants of the major decomposition algorithms and show that a local minima for (2.3a) is also a global minima. (In a related paper Wets (1966b) also shows that the solution set to (2.3a) is both convex and polyhedral.)

Dantzig and Madansky (1960) first proposed applying the Dantzig-Wolfe decomposition algorithm to the dual of a special case of (2.3)

$$
\begin{array}{llllllll}
\text{MIN} & \mathbf{cx} & + p_1\mathbf{d}_1\mathbf{y}_1 & + p_2\mathbf{d}_2\mathbf{y}_2 + & \cdots & + p_S\mathbf{d}_S\mathbf{y}_S & & \\
\text{S.T.} & \mathbf{Ax} & & & & & = & \mathbf{b} \\
& \mathbf{T}_1\mathbf{x} & + \mathbf{W}_1\mathbf{y}_1 & & & & = & \omega_1 \\
& \mathbf{T}_2\mathbf{x} & & + \mathbf{W}_2\mathbf{y}_2 & & & = & \omega_2 \\
& \vdots & & \vdots & & & & \vdots \\
& \vdots & & & \vdots & & & \vdots \\
& \mathbf{T}_S\mathbf{x} & & & & + \mathbf{W}_S\mathbf{y}_S & = & \omega_S \\
& & \mathbf{x} \geq \mathbf{0};\ \mathbf{y}_i \geq \mathbf{0},\ i = 1, \ldots, S & & & & & (2.14)
\end{array}
$$

where there exists a finite number of ω_i ($i = 1, 2, \ldots, s$) with known probability p_i. Wets (1966a) proposes a modification to the dual of (2.14) whereby the normal dual variables π_i are replaced with $\bar{\pi}_i = (1/p_i)\cdot\pi_i$. The subsequent dual set of inequalities corresponding to the column vectors associated with $\mathbf{x}$ in the primal then form the master problem in the dual, while the set of constraints corresponding to decision variable $\mathbf{y}_i$ in (2.14) create subproblem i in the dual. Wets also discusses another version of (2.14) whereby subtracting the $k^{\underline{th}}$ row $\mathbf{T}^k\mathbf{x} + \mathbf{W}^k\mathbf{y}_k = \omega^k$ from the $K{+}1^{\underline{st}}$ row $\mathbf{T}^{k+1}\mathbf{x} + \mathbf{W}^{k+1}\mathbf{y}_{k+1} = \omega^{k+1}$ generates a staircase system with the same constraints as in the recourse section (2.14) (except for $\mathbf{T}^1\mathbf{x} + \mathbf{W}^1\mathbf{y}_1 = \omega^1$), thus providing a simpler form for computation (Wets 1966a). (Also see Wets (1974,1988) for a more recent summary of his results.)

In 1969, Van Slyke and Wets developed the *L*-shaped algorithm based on the immediate result of Wets (1966b) proof that if "...the set of feasible decisions, represented by an n-vector $\mathbf{x}$, is a convex polyhedral subset of R^n...at most a *finite* number of linear constraints must be added to the problem (2.3a) to determine the set of feasible decisions (Van Slyke and Wets 1969)." For such a finite case their algorithm iteratively adds two types of constraints to (2.3a) — both in terms of $\mathbf{x}$ — that: (1) reduce the region of (2.3a) as necessary to guarantee feasibility for (2.3b) (*feasibility cuts*); and, (2) produce an optimal solution to (2.3a) through the dual variables found in solving the recourse problem (2.3b) (*optimality cuts*). Unfortunately, Van Slyke and Wets point out that for the continuous case of ω (including $\mathbf{T}$), developing the second set of constraints requires knowing the entire description of its probability space, and can easily lead to an infinite

number of simplex multiplier-based constraints. Furthermore, they note that even problems with finite distributions of ω may have a tremendously large set of values, thus posing practical computational problems. Finally, they suggest investigating approximation schemes, although such methods may eliminate the optimal solution through sampling error (Van Slyke and Wets 1969). (Wets (1972) also provides characterization theorems and algorithms to test whether a recourse problem is feasible, bounded, and solvable prior to solving it.)

Garstka and Rutenberg (1973) present one of the first enhancements to the *L*-shaped algorithm by developing a more efficient method of calculating the dual variables of the recourse problem associated with the optimality cuts. Taking advantage of the lattice structure of a finite distribution of ω, they perform a parametric ranging of all combinations of the elements (lattice points) in ω regarding their feasibility for a given optimal basis. They justify this approach by noting that the information provides a probability estimate for each realization of ω that in turn helps calculate the dual variables in the *L*-shaped algorithm. Garstka and Rutenberg then incorporate this idea in a basis generation procedure that removes a vector from the current optimal basis and replaces it under the dual simplex procedure. At each optimal basis their algorithm systematically classifies the lattice points according to their feasibility using the above parametric ranging procedure. Garstka and Rutenberg show their *sifting* algorithm to be more efficient than earlier techniques that sequentially try *all* lattice points against an optimal basis, then find another optimal basis from which to try those lattice points that were infeasible under the previous basis, and so on until all realizations of ω are allied with an optimal basis (Garstka and Rutenberg 1973). While this

method can provide faster execution of the *L*-shaped algorithm, it can also iteratively supply finer discrete approximations for continuous cases of ω (Wets 1983). Also Wets (1988) points out that *bunching* procedures may be more appropriate for cases where ω cannot be represented in a lattice structure, contains dependent random variables, or derives from an approximation design.

Both approaches share the assumption that a relatively small number of optimal bases exist for the recourse problem (2.3b) for all realizations of ω - **Tx**. Garstka and Rutenberg (1973) claim that even with a large number of lattice points "...the [recourse] subproblems are very similar indeed, differing only by some minor change to an element in p' [the random right-side vector]. It seems likely that a great many of the...optimal bases...will be the same...so there is the combinatorial problem of spotting them in some systematic manner..." Similarly, Wets (1983) notes that "...because of the nature of the problem at hand it is reasonable to expect that only a small number of bases in W (**W** in (2.3b)) will suffice to bunch all the realizations." Furthermore, in related areas of stochastic programming such as the distribution problem, some numerical algorithms implicitly make the same assumption by parametrically decomposing the sample space of ω into decision regions (optimal bases) for later use with updated probability distributions (Bereanu 1980). One option this dissertation proposes for solving and characterizing the recourse function incorporates this assumption of a few optimal bases in the recourse problem. Chapter 3 explains in detail how the proposed Monte Carlo simulation framework implements this approach for solving recourse problems.

Kall (1979) proposes a *basis factorization* technique (see Strazicky 1974) for exact solution methods by taking advantage of the block structure of the dual of (2.14). He suggests representing the dual basis **B** of any feasible solution to (2.14) in the form

$$\mathbf{B} = \begin{pmatrix} \mathbf{T} & \mathbf{Y} \\ \mathbf{L} & \mathbf{Z} \end{pmatrix} \tag{2.15}$$

where submatrix **T** is regular and invertable. Kall shows how all iterations of the simplex can maintain the form of (2.15) with submatrix **T**; and, through a reformulation of the dual of (2.14) constructs a block-diagonal form for **T** as well. Where **A** is mxn, **W** is uxv, r the number of scenarios or realizations of ω in (2.14), Kall (1979) shows the number of operations per simplex iteration on (2.15) for the factorization method to be on order $O(r)$, where standard pivots on (2.15) are on order $O([r(\mathrm{v} - \mathrm{u}) + n]^2)$. However, Birge (1988) shows that this dual basis factorization requires the same computational effort as the L-shaped method.

Birge and Louveaux (1988) propose an extension of the L-shaped method that provides (under certain conditions) faster convergence to an optimal solution than Van Slyke and Wets' original algorithm. Their algorithm recognizes that in inner linearization (column formation) decomposition algorithms the rate of optimal solution convergence may increase with multiple — as opposed to single — column generation proposed by Birge (1985). Although the L-shaped method performs an outer linearization (row generation) of (2.3a) through the dual variables of the recourse problem, Birge and Louveaux hypothesize that the same phenomenon can occur in such outer linearization methods. Consequently, their

extension generates multiple optimality cuts at the same iterative point where Van Slyke and Wets' algorithm generates one. Furthermore, Birge and Louveaux establish an upper bound on the number of iterations for the two versions based on the maximum number of finite convex sets of the recourse polyhedron *(b)*, the number of constraints in the recourse problem *(m_2)*, and the number of realizations of ω (K). These bounds — $1+ K[b^{m2} -1]$ for their multi-cut method versus $(1 + K[b - 1])^{m2}$ for the L-shaped algorithm — shows how the multi-cut method could have an advantage for large m_2 (Birge and Louveaux 1988).

According to Helgason and Wallace (1991) the L-shaped method dominated most computational research in stochastic programming until Wets (1989) and Rockafeller and Wets (1991) proposed another decomposition method called *scenario aggregation*. As explained by Wets (1989), scenario aggregation differs from the earlier decomposition algorithms by its underlying assumption that the random elements of the problem *cannot* be accurately described by a probability distribution. Instead of solving the problem through discrete approximations based on distributional assumptions, Wets describes this lack of information leading to *scenario analysis*; i.e., characterizing the randomness through a relatively few scenarios. Specifically, Wets proposes an algorithm that finds a solution $\mathbf{x}^*$ to the problem

$$\begin{aligned} \text{MIN} \quad & \Sigma\, p_s f(\mathbf{x}, s^N) \\ \text{s.t.} \quad & \mathbf{x} \in \cap_s C_s \end{aligned} \tag{2.16}$$

where $N = \{s^1, \ldots, s^L\}$ represents the set of scenarios s^i and p_s is the probability weight, C_s the set of solutions, and $f(\mathbf{x},s^i)$ is the objective function for scenario s^i,

respectively. Letting $\mathbf{x}^{vs}$ represent the optimal solution to the individual scenario MIN $f^v(\mathbf{x},s)$ s.t. $\mathbf{x} \in C_s$ at iteration v, Wets progressively updates $\mathbf{x}^{vs}$ with an average solution $\hat{\mathbf{x}}^v = \Sigma p_s \mathbf{x}^{vs}$ and

$$f^v(x,s) = f(x,s) + w^{v-1}(s)\mathbf{x} + \frac{1}{2}\rho|\mathbf{x} - \hat{\mathbf{x}}^{v-1}|^2 \tag{2.17}$$

where

$$w^v(s) = w^{v-1}(s) + \rho|\mathbf{x}^{vs} - \hat{\mathbf{x}}^v|. \tag{2.18}$$

Defining an *implementable* solution as one where $\hat{\mathbf{x}}$ is scenario independent and an *admissible* solution to be one where $\hat{\mathbf{x}}$ is feasible for all C_s, he shows that an optimal solution to (2.16) meets both conditions when $\mathbf{x}^{vs} = \hat{\mathbf{x}}^v$ for all $s \in N$. Wets also argues that unlike a simple averaging of the optimal solutions $\mathbf{x}^s$ to the individual scenario problems $f(x,s)$, a solution to (2.16) allows for the costs that occur for choosing a given $\mathbf{x}$ and realizing scenario s (Wets 1989). Although not directly related to this dissertation's area of research (two-stage problem with recourse), it should be noted that Wets (1989) and Rockafeller and Wets (1991) extend the scenario aggregation policy to multiple periods by modifying the method to calculate a solution set conditioned upon known or observable information about the problem in each period. This scenario aggregation technique that produces an implementable and admissible solution under such non-anticipitivity conditions for multi-period problems is known as the *progressive hedging* algorithm (Wets 1989, Rockafeller and Wets 1991). Recent applications of the progressive hedging algorithm include Lagrangian

approximations of the individual scenarios (Helgason and Wallace 1991) and stochastic network programming (Mulvey and Vladimirou 1991, 1992).

2.3.3 Sampling Methods

The previous sections discussed solution techniques for (1) cases where realizations of the random vector ω are exact, known, and of manageable size; and, (2) approximation methods such as aggregation or partitioning for distributions of ω that are either continuous, possess large discrete realizations, or are only partially known. The literature offers another approach using sampling techniques under two basic contexts — decomposition and stochastic quasigradient (SQG) algorithms. In general, both aggregation and stochastic decomposition involve algorithms that work with only a small fraction of the overall sample space, thus implying a need to determine solution quality (Higle and Sen 1993). Although the SQG approach primarily employs gradient search techniques, it too only has a small exposure to the sample space that causes problems in estimating the solution quality. The literature on SQG methods is covered in Section 2.3.4.

Higle and Sen (1991b) propose a variant of the decomposition approach for two-stage recourse problems based on sampling the random vector ω of subproblem (2.3b). For a given feasible solution $\mathbf{x}^k$ to (2.3a), Higle and Sen observe that the dual of (2.3b) is

$$\begin{aligned} h(\mathbf{x}^k,\omega) = \quad & \text{MAX} \quad \pi^k(\omega - \mathbf{T}\mathbf{x}^k) \\ & \text{s.t.} \quad \pi^k\mathbf{W} \le \mathbf{d} \qquad \pi^k \text{ unrestricted} \end{aligned} \tag{2.19}$$

and where they define V as the set of all vertices for the dual, V^k as the set of dual vertices $\{\pi^1,\pi^2, \ldots ,\pi^k\}$, and ω^t an independent sample of ω for $t = 1, \ldots , k$, then

$$\text{MAX}\{\pi(\omega^t - \mathbf{T}\mathbf{x}^k) \mid \pi \in V^k\} \leq \text{MAX}\{\pi(\omega^t - \mathbf{T}\mathbf{x}^k) \mid \pi \in V\} \qquad (2.20)$$

and

$$\pi^k_t(\omega^t - \mathbf{T}\mathbf{x}^k) \leq h(\mathbf{x}^k,\omega^t). \qquad (2.21)$$

Higle and Sen also define a piecewise linear approximation of (2.3b) after k iterations as

$$f_k(\mathbf{x}) = \text{MAX}\{\alpha^k_t + (\beta^k_t + \mathbf{c})\mathbf{x} \mid t = 1, \ldots,k\} \qquad (2.22)$$

which in turn forms the master program

$$\text{MIN}\, f_k(\mathbf{x}), \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0} \qquad (2.23)$$

Based on the results of (2.20-21), their algorithm at the k^{th} step receives an optimal solution $\mathbf{x}^k$ from (2.23) (based upon k-1 previous cuts) and generates a new cut k of the form

$$\alpha^k_k + (\beta^k_k + \mathbf{c})\mathbf{x} \equiv \mathbf{c}\mathbf{x} + \frac{1}{k}\sum_{t=1}^{k} \pi^k_t(\omega^t - \mathbf{T}\mathbf{x}) \qquad (2.24)$$

for inclusion in (2.22) at the k+1 iteration. Additionally, they update the previous cuts using the formula

$$\alpha^k_t = \frac{k\text{-}1}{k}\alpha^{k-1}_t, \quad \beta^k_t = \frac{k\text{-}1}{k}\beta^{k-1}_t, \quad t = 1, \ldots , k - 1 \qquad (2.25)$$

in order to provide *all* generated cuts with current sampling information. Higle and Sen (1991b) essentially build an approximate outer linearization of (2.3a) using cuts generated by (2.24) and independent samplings of ω and $\mathbf{T}$; in this fashion, they avoid any requirement for finite realizations of ω and $\mathbf{T}$ (or equivalent discrete approximations) and any distributional assumptions on their random components.

Higle and Sen (1991b) also address the convergence issue their algorithm poses due to the randomized nature of the cuts generated in (2.24) and (2.25) by proposing a modification to their basic algorithm using an incumbent solution $\tilde{\mathbf{x}}^{k-1}$ at the k^{th} iteration. Designating this algorithm *stochastic decomposition*, they use the incumbent solutions $\tilde{\mathbf{x}}^{k-1}$ and $\tilde{\pi}^{k-1}$ to update all previous cuts (2.25) with the current realization ω^t, then substitute the incumbent solution $\tilde{\mathbf{x}}^{k-1}$ with $\mathbf{x}^k$ if the condition

$$f_k(\mathbf{x}^k) - {}_k(\tilde{\mathbf{x}}^{k-1}) \le r\cdot\{f_{k-1}(\mathbf{x}^k) - f_{k-1}(\tilde{\mathbf{x}}^{k-1})\}, \tag{2.26}$$

where r is fixed on the interval [0,1], is met. Higle and Sen also show that at least one of the test statistics

$$\gamma^k = \frac{1}{k}\sum_{i=1}^{k} f_i(\tilde{\mathbf{x}}^i) \qquad \text{or} \qquad \tilde{\gamma}^k = \frac{1}{m_k}\sum_{i=1}^{mk} f_{kn}(\tilde{\mathbf{x}}^{kn}), \tag{2.27}$$

where m_k represents the number of incumbent solutions detected by iteration k, converges to the optimal solution value, thus suggesting a stopping criteria of the form

$$\frac{|f_k(\tilde{\mathbf{x}}^k)-\gamma^{k-1}|}{\gamma^{k-1}} < \varepsilon \qquad \text{or} \qquad \frac{|f_k(\tilde{\mathbf{x}}^k)-\tilde{\gamma}^{k-1}|}{\tilde{\gamma}^{k-1}} < \varepsilon, \tag{2.28}$$

respectively. They also suggest comparing the cardinality of the sets of vertices V^k and V^{k+1} to the dual (2.19) as a method to prevent premature termination (Higle and Sen 1991b).

In a separate article, Higle and Sen (1991a) offer additional termination criteria to the stochastic decomposition algorithm based upon the empirical results of their algorithm arriving at an optimal solution prior to solution stability as represented in (2.28). They propose alternative termination methods by either resampling the observations on ω (using Lagrangian duality and Kuhn-Tucker conditions) or resampling the dual vertices of (2.19) (Kuhn-Tucker alone). Their results from an example power expansion program found that the stopping point for the three proposed statistical verifications of optimality occurred at fewer iterations than the original objective function stability measure (2.28). Finally, Sen, Mai, and Higle (1994) provide a summary of the stochastic decomposition algorithm within the framework of a randomized versions of Kelly's (1960) cutting plane methods and Benders' (1962) decomposition of two-stage linear programs.

Dantzig and Glynn (1990) propose a method that combines Benders' decomposition, Monte Carlo sampling, and parallel processing to solve multi-period problems. Their idea places the master problem under the control of a single processor that iteratively provides updated solutions $\mathbf{x}$ to several parallel computers responsible for solving the dual subproblems for sample realizations of

ω. The subprocessors in turn update the master program with cuts used to develop a new estimate of $\mathbf{x}$. They terminate the algorithm when the difference between the upper bound for MIN $Z(\mathbf{x})$ and the lower bound (cuts from the subproblems) reaches a pre-specified interval. For continuous or large finite cases of ω, Dantzig and Glynn suggest a Monte Carlo sampling procedure for the subproblem; and, to improve both the convergence (through variance reduction) and robustness of the algorithm they incorporate importance sampling of the scenarios (Dantzig and Glynn 1990. Also see Dantzig and Infanger 1991). Sen, Mai, and Higle's (1994) comparison of Dantzig and Glynn's approach to stochastic decomposition notes that the former suffers from: (1) using a fixed, independent samplings that detract from calculating the error bounds; and (2) solving one subproblem for each sample ω at each iteration versus no more than 2 subproblems per iteration for the stochastic decomposition method.

Berger, Mulvey, and Ruszczynski (1993) propose another method for solving scenario-based models based upon the idea of dualizing the non-anticipitivity constraints (i.e., the first-stage decision variables $\mathbf{x}$ in (2.14)) to produce separable programming problems for each scenario. For example, (2.14) can be reformulated to read

$$
\begin{array}{lrcl}
\text{MIN} & \mathbf{c}\mathbf{x}_1 + p_1\mathbf{d}_1\mathbf{y}_1 \quad +\ldots+ \quad \mathbf{c}\mathbf{x}_S + p_S\mathbf{d}_S\mathbf{y}_S & & \\
\text{s.t.} & \mathbf{A}\mathbf{x}_1 & = & \mathbf{b} \\
& \mathbf{T}_1\mathbf{x}_1 + \mathbf{W}_1\mathbf{y}_1 & = & \omega_1 \\
& \vdots & & \vdots \\
& \mathbf{A}\mathbf{x}_S & = & \mathbf{b} \\
& \mathbf{T}_S\mathbf{x}_S + \mathbf{W}_S\mathbf{y}_S & = & \omega_S \qquad (2.29a)
\end{array}
$$

$$\begin{aligned} \mathbf{x}_1 - \mathbf{x}_2 &= \mathbf{0} \\ \vdots & \quad \vdots \\ \mathbf{x}_{S-1} - \mathbf{x}_S &= \mathbf{0} \end{aligned}$$

$$\mathbf{x}_i \geq \mathbf{0};\ \mathbf{y}_i \geq \mathbf{0}. \tag{2.29b}$$

Redefining the notation of (2.29) to where

$$\mathbf{A}_i = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{T} & \mathbf{W}_i \end{pmatrix} \quad \mathbf{b}_i = \begin{pmatrix} \mathbf{b} \\ \omega_i \end{pmatrix} \quad \mathbf{c}_i = \begin{pmatrix} \mathbf{c} \\ \mathbf{d}_i \end{pmatrix} \quad \mathbf{x}_i = \begin{pmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{pmatrix} \tag{2.30}$$

they let $\mathbf{N}_i$ be the i^{th} period where element (j,k) of N equals 1 if and only if $j = k$ and the j^{th} component of N shares the same history through period i+1. Under these conditions, a general multi-period recourse formulation (where the two-stage recourse problem is a special case) can be written as

$$\text{MIN} \quad \sum_{i=1}^{S} \mathbf{c}_i\mathbf{x}_i \tag{2.31a}$$

$$\text{s.t.} \quad \mathbf{A}_i\mathbf{x}_i = \mathbf{b}_i \qquad i = 1, \ldots, n \tag{2.31b}$$

$$\mathbf{N}_i\mathbf{x}_i - \mathbf{N}_i\mathbf{x}_{i+1} = \mathbf{0} \qquad i = 1, \ldots, n-1 \tag{2.31c}$$

$$\mathbf{x}_i \geq \mathbf{0} \qquad i = 1, \ldots, n. \tag{2.31d}$$

Berger, Mulvey, and Ruszczynski then dualize constraint (2.31b) by dropping (2.31c) and replacing (2.31a) with

$$\sum_{i=1}^{S} \mathbf{c}_i\mathbf{x}_i + \sum_{i=1}^{S} \pi_i(\mathbf{N}_i\mathbf{x}_i - \mathbf{N}_i\mathbf{x}_{i+1}) + \frac{1}{2} r \sum_{i=1}^{S} \|\mathbf{N}_i\mathbf{x}_i - \mathbf{N}_i\mathbf{x}_{i+1})\|^2. \tag{2.32}$$

They propose an approximation method for solving the non-separable quadratic form of (2.32) using a nonlinear interior point method whereby the approximate solution $\hat{\mathbf{x}}$ and the multipliers π_i are updated every 2-4 and 100-150 iterations, respectively (Berger, Mulvey, and Ruszczynski 1993). Berger and Mulvey (1993) propose an improvement of this algorithm with a restart strategy that addresses the instability of interior point codes after updating $\hat{\mathbf{x}}$.

2.3.4 Subgradient Methods

Sampling methods for recourse problems using probability distributions to model the random parameters are found in the literature in two basic categories: (1) the sampling-based algorithms applied in a decomposition context (such as Higle and Sen's stochastic decomposition algorithm) just reviewed; and (2) stochastic quasigradient (SQG) algorithms (Morton 1994a). The SQG method attempts to find the solution to the recourse problem by extending the classic steepest descent (or gradient search) method found in non-linear programming to the stochastic programming environment. As explained by Luenberger (1989), in a deterministic setting we wish to find

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha^k \nabla f(\mathbf{x}^k) \tag{2.33}$$

where $\nabla f(\mathbf{x}^k)$ is the gradient of the function $f(\mathbf{x})$ defined as

$$\nabla f(\mathbf{x}^k) = \left[\frac{\partial f(\mathbf{x})}{\partial x_1} \quad \frac{\partial f(\mathbf{x})}{\partial x_2} \quad \cdots \quad \frac{\partial f(\mathbf{x})}{\partial x_M} \right] \tag{2.34}$$

and α^k is a scalar that minimizes $f(\mathbf{x}^k - \alpha^k \nabla f(\mathbf{x}^k))$. If $\mathbf{x} \in \boldsymbol{X}$ constrains (2.33), then a *projection gradient* method determines the descent direction by projecting the gradient onto the working surface to maintain feasibility (Luenberger 1989). As Law and Kelton (1991) point out, (2.34) obviously cannot be directly applied in the stochastic environment due to random variation of $f(\mathbf{x})$; hence, a gradient estimation technique (such as *replication* or *perturbation analysis*) must be employed. SQG algorithms extend these stochastic approximation algorithms to stochastic programming problems (Ermoliev 1983).

Defining $F(\mathbf{x}) = E_\omega[f(\mathbf{x},\omega)]$ as the expected value of the objective function, Gaivoronski (1988) provides a general extension of (2.33) for the stochastic programming problem as

$$\begin{aligned} \mathbf{x}^{k+1} &= \pi_{\mathbf{x}}[\mathbf{x}^k - \alpha^k \nu^k],\ k = 0, 1, \ldots \\ E[\nu^k \mid \mathbf{x}^0, \mathbf{x}^1, \ldots, \mathbf{x}^k] &= F_{\mathbf{x}}(\mathbf{x}^k) + b^k \end{aligned} \tag{2.35}$$

where $\pi_{\mathbf{x}}$ is the projection operator, $\mathbf{x}^k$ is the incumbent solution approximation, α^k is the step size, and ν^k (as a statistical estimate of the subgradient of $F(\mathbf{x})$) is the *stochastic quasigradient* of $F(\mathbf{x})$. In general, he notes that the quasigradient ν^k can be estimated as

$$\nu^k = \frac{1}{N} \sum_{i=1}^{N} \nabla f(\mathbf{x}^k, \omega^i) \tag{2.36}$$

using Monte Carlo sampling (with asymptotic error $1/\sqrt{N}$), or where distributional information on ω is available he suggests potential asymptotic error rates approaching $\log(N)/N$. He also observes that SQG methods do not exhibit

monotonicity due to the random nature of (2.36), thus contributing to problems in determining optimality (i.e., $\|\mathbf{x}^{k+1} - \mathbf{x}^k\| \leq \varepsilon$), calculating step size α^k, and estimating step direction ν^k. Finally, Gaivoronski reviews several alternatives to (2.36) such as random search analogs, finite difference approximations

$$\nu^k = \sum_{i=1}^{N} \frac{f(\mathbf{x}^k+\delta^k e_i,\omega^k_i)-f(\mathbf{x}^k+\omega^k_i)}{\delta^k} \tag{2.37}$$

or central finite differences

$$\nu^k = \sum_{i=1}^{N} \frac{f(\mathbf{x}^k+\delta^k e_i,\omega^k_i)-f(\mathbf{x}^k-\delta^k e_i,\omega^k_i)}{2\delta^k} \tag{2.38}$$

where e_i is the unit vector and δ^k is a scalar; or, in the case where $F(\mathbf{x})$ is not differentiable he suggests objective function smoothing, or averaging using

$$\nu^{k+1} = \frac{1}{M} \sum_{i=k-M+1}^{k} \nu^k \tag{2.39}$$

where M is memory size (Gaivoronski 1988). In the case of the recourse function Ermoliev (1983, 1988) shows that for a single sample ω^S

$$\nu^k = \mathbf{c} + \mu(\mathbf{x}^k,\omega^S)\mathbf{T} \tag{2.40}$$

where $\mu(\mathbf{x}^k,\xi^S)$ are the dual variables for the recourse problem.

Recent research addresses the major drawbacks of SQG techniques; i.e., slow convergence, oscillation in the neighborhood of the optimal solution (stopping times), projection on $\boldsymbol{X}$, and selecting the appropriate stepsize.

Ruszczynski and Syski (1986) propose an auxiliary filter that provides aggregate stochastic subgradients at each iteration. Pflug (1988) reviews both of these problems in the context of *deterministic*, *adaptive*, *ratio-of-progress*, *oscillation*, and *inner product* tests, as well as providing comparison and implementation comments. Urasiev (1988) discusses an adaptive procedure using quasigradient directional information to calculate subsequent stepsizes, while Marti (1988) proposes a semi-stochastic approximation that under certain cases can partially restore monotonicity to the objective function. Rockafeller and Wets (1988) give a method for gradient projection on $\boldsymbol{X}$ that, under certain conditions, does not require penalization or primal-dual workarounds. In addition to Gaivoronski (1988), Ermoliev (1988), and Ermoliev and Nurminisky (1980) provide general introductions to SQG methods.

2.4 Simulation and SLP with Recourse

Referring to Figure 1.1, the previous sections of this chapter summarize an extensive amount of research in the categories of *Search Techniques* and *Optimization Algorithms*. In all cases the research objective attempts to provide a better method for finding the optimal solution, and in many cases does so by improving the statistical estimation of the stochastic effects within a mathematical programming framework. *However, an extensive review of the literature found no attempts to model stochastic two-stage linear programming programs with fixed and complete recourse from a simulation perspective.* Consequently, very little has been accomplished in investigating the remaining four categories of *Variance Reduction*, *Experimental Design*, *Response Surface Analysis*, and

Distributional Analysis since these topics naturally arise in a simulation environment. (The efforts by Dantzig and Glynn (1990), Danztig and Infanger (1991), and Infanger (1994) in applying *Importance Sampling* to stochastic programming provide one exception in the open literature for using variance reduction methods.) This absence of research thus provides this dissertation with its principal focus and contributions, and constitutes the subject matter of Chapters 4 and Chapter 5. Therefore, this section will offer only a general overview of simulation optimization, and will defer the application of these techniques in the SLP context to the following chapters.

Azadivar (1992) provides a brief overview of the four major approaches to using simulation as an optimization tool.

1. *Gradient-Based Search Methods* derive from traditional non-linear programming; the most-often used methods include finite difference estimation, infinitesimal perturbation analysis, frequency domain analysis, and likelihood ratio estimators.

2. *Stochastic Approximation Methods* apply recursive search techniques that converge on the theoretical minimum or maximum.

3. *Response Surface Methodology* fits a low-order polynomial approximation to the responses of the simulation, usually derived from a formal experimental design.

4. *Heuristic Methods* fall into two broad categories — complex search and simulated annealing (Azadivar 1992).

Within this construct the algorithms described in Chapter 3 incorporate items (1) and (2) by employing a recursive search technique referred to as the Nelder-Mead simplex method (Nelder and Mead 1965), and the projected gradient and PARTAN search methods discussed in Luenberger (1984). Chapter 4 develops the techniques for fitting and analyzing a response surface to $Z(\mathbf{x})$ in (2.3a) for a given solution $\mathbf{x}$, while item (4) is not used in this thesis.

Simulation does have some disadvantages and pitfalls. Summarizing Law and Kelton (1991), these problems include:

1. *Expense.* Simulation models of complex systems can be costly and time-consuming to develop and run.

2. *Stochastic Nature.* Although inherently stochastic, simulation models perform better when comparing alternative solutions rather than finding the optimal one.

3. *Bad Assumptions.* While false assumptions can derail any modeling effort, simulation is especially vulnerable to mistaken probability distributions, false presumptions of independence, inaccurate identification of randomness, biased random number generation, and an insufficient number of replications.

4. *Resolution.* Getting the right level of detail is important — and difficult. If there is too little detail, then the model lacks validity; too much, and it introduces extraneous noise to the response, becomes expensive to code, and time-consuming to run (Law and Kelton 1991).

The process of calculating the response surfaces associated with (2.3a) is perhaps most susceptible to the drawbacks of expense and bad assumptions. Clearly, using a repetitive search will be computationally more expensive than current methods; however, this dissertation contends (and the research will show) that the additional information justifies the effort. As a simulation, the proposed methodology is also vulnerable to bad assumptions, although this drawback applies to the most recent LP-oriented approaches (e.g., stochastic decomposition) as well. The research assumes that the probability distributions for the test problems in Chapter 5 are correct, and will address the issues of bias (through either random number generation or other sources) and insufficient replications at the appropriate point.

Additional information and details on simulation optimization techniques can be found in Barton and Ivey (1991), Biles and Swain (1979), Box and Draper (1987), Fu (1994), Jacobson and Schruben (1989), Kleijnen (1974, 1987), Law and Kelton (1991), Meketon (1987), Pritsker (1986), Rubinstein (1981), and Safizadeh (1990).

Chapter 3

Methodology: Optimization

3.1 INTRODUCTION

3.1.1 Overview

This chapter describes the optimization methodology for deriving the response surface approximation of the expected optimal value of the objective function associated with a class of two-stage stochastic linear programming problems with relatively complete and fixed recourse. The following sections of this introduction review the notational form, principal terms, and major definitions for the remainder of the dissertation; provide an exact description of the class of recourse problems investigated by this research; and, present a sample recourse problem used throughout Chapters 3 and 4 to illustrate the particular techniques. After this introduction, Sections 3.2 and 3.3 present, respectively, two major areas regarding optimality and algorithmic efficiency investigated by this research: *Non-Linear Search Techniques* and *Linear Programming Algorithms*. The next chapter covers the statistical analysis topics of *Variance Reduction*, *Experimental Design*, *Canonical Analysis*, and *Distribution Analysis*. Although this chapter focuses on the computational aspects of deriving a response surface approximation of the expected value of a two-stage LP with recourse, when appropriate it also provides the theoretical background, literature references, and the author's own contributions for the area under study.

3.1.2 Model Description and Basic Definitions

For notational purposes matrices are shown in uppercase boldface with the subscript identifying the specific matrix (vectors appear in lowercase boldface). Where a matrix is identified in boldface type, the superscripts label column or row vectors; conversely, where an element of a matrix appears in non-boldface type the superscripts describe the location of the element in the matrix. Any additional subscripts appearing in parenthesis refer to that matrix as defined for the equation in the subscripted parenthesis. If no subscripted item in parenthesis appears, then assume the matrix or element in the context of its most recent definition. The superscript "*" identifies that matrix as associated with an optimal solution for an optimization problem. Subscripting conventions also apply to functional notation.

> **Example**. $\mathbf{G}^{k\bullet}_{i(2)}$ refers to the k^{th} row in the i^{th} version of matrix **G** as that matrix is defined in the set of equations (2).

> **Example**. G^{ij} refers to the element in the i^{th} row and j^{th} column of matrix **G** as most recently defined.

This dissertation restricts its research to a specific, but important, category of two-stage programming under uncertainty with relatively complete and fixed recourse — the *capacity expansion problem*. In general, the capacity expansion problem involves optimization through a first-stage decision (**x**) regarding the amount of production capacity to add, with a follow-up second-stage vector (**y**) typically deciding the optimal resource allocation *after* the realization of any random variables. Examples of this sort of problem from the literature and

INTERNET include power expansion, machine capacity expansion, and facility location problems; and, almost always contains uncertainty in one or more resource's availability or requirement (i.e., demand, budget, physical limitations). Although the second-stage production problem possesses an infinite horizon (i.e., periodic and recurring), most models assume (with appropriate present-value adjustments in the objective function coefficients) that a simplifying single second-stage model captures the essential behavior of a more complicated multi-stage expression of the recourse problem. Also, expansion problems can involve a sequence of *capacity decisions over time*, but this research restricts its focus to a one-time expansion decision formulated in the first-stage.

Mathematically, the two-stage capacity expansion problem examined by this research is expressed as MIN $Z(\mathbf{x})$ where

$$Z(\mathbf{x}) = \mathbf{cx} + E[h(\mathbf{x},\omega,\mathbf{T})], \text{ s.t. } \mathbf{Ax} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0} \tag{3.1a}$$

$$h(\mathbf{x},\omega,\mathbf{T}) = \text{MIN } \mathbf{dy}, \text{ s.t. } \mathbf{Wy} = \omega - \mathbf{Tx}, \ \mathbf{y} \geq \mathbf{0} \tag{3.1b}$$

c and **d** are cost coefficient vectors for a unit increase in **x** and the recourse decision **y**, respectively; **A** is the matrix of per unit consumption of resource **b** by **x**; **W** is the matrix of per unit consumption of resource ω as adjusted by the vector **Tx**; and function $h(\mathbf{x},\omega,\mathbf{T})$ is defined as the *recourse problem* (3.1b). Alternatively, (3.1a) can be expressed in a profit maximization form, where **d** represents the profit gained as offset by the cost of expansion in **c**; however, throughout this dissertation the text assumes a minimization objective. Additionally, all formulations assume that a finite optimal solution exists for (3.1a). Although problem specific, this condition can be explicitly guaranteed by

including a set of error vectors to formally insure feasibility, yet indicating through their presence in any optimal bases of constraint violations (see Mulvey 1993). While uncertainty can also occur in objective function and constraint coefficients, this dissertation follows the recent literature in restricting random variation to the right-side vector ω - $\mathbf{Tx}$. In terms of (3.1a) and (3.1b), all coefficients possess a fixed value except ω and $\mathbf{T}$, where the model assumes that a finite mean and variance exists for each element. Finally, *no assumption is made on the independence of the random components of* ω *and* $\boldsymbol{T}$.

The literature provides two important characteristics of $Z(\mathbf{x})$: (1) $Z(\mathbf{x})$ is a convex linear function of $\mathbf{x}$ (piecewise convex linear function for finite realizations of ω and $\mathbf{T}$); and, (2) *the only true decision vector is* $\boldsymbol{x}$, since $\mathbf{y}$ deterministically follows the decision vector $\mathbf{x}$ and the realization of the random variables in ω and $\mathbf{T}$. Consequently, a second-order polynomial approximation can be fit to the true function $Z(\mathbf{x})$ in the region of optimality (defined shortly) using estimates derived from a Monte Carlo simulation. Indeed, in theory during the course of the simulation of $Z(\mathbf{x})$ several responses to $\mathbf{x}$ could be approximated:

1. Point estimates (and associated upper and lower confidence intervals) of the first- and second-moments of the distribution describing $h(\mathbf{x},\omega,\mathbf{T})$.

2. The relative frequency of different bases of the recourse problem $h(\mathbf{x},\omega,\mathbf{T})$ being optimal.

3. By extension of item (2), the estimated values of $\mathbf{y}$ for (3.1b).

However, for several reasons this research will fit a response surface only to $Z(\mathbf{x})$. First, the primary focus of this work is to establish — *for the first time* — the principle and techniques of fitting a response surface to an important aspect of (3.1a). Second, this inquiry presents search techniques that rely heavily on the convexity property of $Z(\mathbf{x})$; conversely, such characteristics for any other response has yet to be shown. Finally, the dynamic and generally unknown nature of the underlying distribution of $h(\mathbf{x},\omega,\mathbf{T})$ strongly suggests proceeding with a non-parametric investigation of the region of optimality rather than trying to fit a response to higher-order moments (Wilson 1995). Therefore, this dissertation leaves other single — or multiple — response estimations for future study.

The following definitions represent the primary terminology used throughout this dissertation:

> **Definition**. Let $m_{(3.1a)}$ and $n_{(3.1a)}$ represent the number of constraints and variables, respectively, for (3.1a). Similarly, let $m_{(3.1b)}$ and $n_{(3.1b)}$ denote the row and column dimensions for (3.1b).
>
> **Definition**. Let $X = \{\mathbf{x} : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$.
>
> **Definition** Let $\mathbf{x}^*$ represent the optimal solution that minimizes $Z(\mathbf{x})$; i.e., $Z(\mathbf{x}^*) = \text{MIN}\ Z(\mathbf{x})$. Define *region of optimality* as the set $\{\mathbf{x}' : Z(\mathbf{x}') \leq Z(\mathbf{x}^*) + \varepsilon, \varepsilon > 0, \mathbf{x}' \in X\}$; i.e. those feasible solutions $\mathbf{x}'$ whose objective values $Z(\mathbf{x}')$ are near-optimal (as defined by ε).

Definition. Let $\mathbf{x}_{ev}$ represent the optimal solution for the expected value approximation MIN$\{\mathbf{cx} + h(\mathbf{x},\mathrm{E}[\omega],\mathrm{E}[\mathbf{T}])\}$, s.t. $\mathbf{x} \in X$. (Recall (2.9) in Section 2.3.1.)

Definition. Let k be an identifying variable for the iterative sequence of first-stage vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_K$ as determined by a search algorithm described in Section 3.2, and let K denote the total number of distinct $\mathbf{x}$ in the sequence where $\mathbf{x}_K = \mathbf{x}^*$ or $\mathbf{x}_K = \mathbf{x}'$.

Definition. Let $z_{ik} = \mathbf{c}\mathbf{x}_k + h(\mathbf{x}_k,\omega_i,\mathbf{T}_i)$; i.e., z_{ik} represents the objective value given $\mathbf{x}_k$ for the i^{th} realization of ω and $\mathbf{T}$.

Definition. Let $\hat{Z}_s(\mathbf{x})$ represent an unbiased estimator of $Z(\mathbf{x})$ using sampling technique s. $\hat{Z}_s(\mathbf{x})$ will be substituted for $Z(\mathbf{x})$ in problems where calculating $Z(\mathbf{x})$ is computationally prohibitive.

3.1.3 Example Problem (APL1P)

The recourse problem and formulation in Figure 3.1 (provided by Morton (1995) and Infanger (1994)) describes the power generation problem APL1P, where the first-stage decision consists of 2 variables x^i (i = 1, 2) that model the supply capacity of their respective source nodes 1 and 2 (the constraints associated with the lower bounds $x^i \geq 1$ correspond to the $\mathbf{Ax} = \mathbf{b}$ portion in 3.1a). The ω^j reflects the stochastic demand of the destination node j (the ω portion in (3.1b)), while ξ^{11} and ξ^{22} represent variation in supply availability (these elements correspond to the $\mathbf{T}$ matrix). Finally, y^{ij} represents the second-stage (recourse)

$$\text{MIN}\quad 4x^1 + 2.5x^2 + 4.3y^{11} + 2y^{12} + .5y^{13} + 8.7y^{21} + 4y^{22} + y^{23} + 10y^{31} + 10y^{32} + 10y^{33}$$

S.T.

$$\begin{array}{|cc|ccccccccccccc|cc|}
\hline
x^1 & & & & & & & & & & & & & \geq & 1 \\
 & x^2 & & & & & & & & & & & & \geq & 1 \\
 & A & & & & & & & & & & & & & \\
\hline
 & T & W & & & & & & & & & & & & \\
-\xi^1 x^1 & & + y^{11} & + y^{12} & + y^{13} & & & & & & & & & \leq & 0 \\
 & -\xi^2 x^2 & & & & + y^{21} & + y^{22} & + y^{23} & & & & & & \leq & 0 \\
 & & + y^{11} & & & + y^{21} & & & + y^{31} & & & & & \geq & \omega^1 \\
 & & & + y^{12} & & & + y^{22} & & & + y^{32} & & & & \geq & \omega^2 \\
 & & & & + y^{13} & & & + y^{23} & & & + y^{33} & & & \geq & \omega^3 \\
\hline
\end{array}$$

T coefficients ξ^i are discrete distributions with the following parameters:

$\text{Prob}\{\xi^1 = 1.0\} = .20$ $\text{Prob}\{\xi^1 = 0.9\} = .30$ $\text{Prob}\{\xi^2 = 1.0\} = .10$ $\text{Prob}\{\xi^2 = 0.9\} = .20$

$\text{Prob}\{\xi^1 = 0.5\} = .40$ $\text{Prob}\{\xi^1 = 0.1\} = .10$ $\text{Prob}\{\xi^2 = 0.7\} = .50$ $\text{Prob}\{\xi^2 = 0.1\} = .10$

$\text{Prob}\{\xi^2 = 0.0\} = .10$

Demand segments ω_j ($j = 1, \ldots, 3$) are discrete distributions with the following parameters:

$\text{Prob}\{\omega^j = 900\} = .15$ $\text{Prob}\{\omega^j = 1000\} = .45$

$\text{Prob}\{\omega^j = 1100\} = .25$ $\text{Prob}\{\omega^j = 1200\} = .15$

FIGURE 3.1
FORMULATION OF APL1P PROBLEM

decision variables that minimize the cost of meeting the demand given the capacity of supply x^i and realization of the stochastic demand ω^j (their technological coefficients constitute **W** and assume a *transportation problem* structure). Representing a standard characteristic of these problems, the recourse decision variables $\mathbf{y}^{3j}$ model the options of purchasing supply outside of the system's capacity as reflected by the higher unit costs $\mathbf{d}^{3j}$; plus, with no upper bound they guarantee a feasible solution for any value **x** and realization of ω. Given its small dimensionality, APL1P provides an excellent case study for graphically demonstrating the proposed techniques for fitting and analyzing a response surface to Z(**x**). Therefore, Chapters 3 and 4 will refer to this problem as needed for illustration.

3.2 NON-LINEAR SEARCH TECHNIQUES

3.2.1 Introduction

The objective of deriving a low-order polynomial approximation to the estimated response Z(**x**) clearly requires experimental data on how Z(**x**) responds to changes in **x**. Formal experimental designs provide the best method for estimating such an approximation (Box and Draper 1987), but ultimately they require knowing where to center the experimental design, which independent factors to include, what levels or values to set them at, and what type of design structure to use. Furthermore, the known convexity of Z(**x**) notwithstanding, the potentially enormous size of X dictates that a quadratic approximation will best fit only a small portion of the sample space (with the evident area of interest being

the region of $\mathbf{x}^*$). Finally, due to its $h(\mathbf{x},\omega,\mathbf{T})$ component $Z(\mathbf{x})$ itself must usually be estimated except for relatively small and discrete distributions of ω and $\mathbf{T}$. Obviously, such experimental design information about any recourse problem of the form (3.1) is not known beforehand and must therefore be found. Consequently, this dissertation proposes adapting several search methods from the simulation and non-linear programming literature as techniques for producing data in a way that allows the construction of a formal experimental design within the region of optimality, and thus deriving an accurate response surface approximation.

The fundamental idea of all the proposed methods involves searching $\boldsymbol{X}$ using an iterative sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$, where $Z(\mathbf{x}_1) \geq Z(\mathbf{x}_2) \geq \ldots \geq Z(\mathbf{x}_k)$, for two basic purposes:

1. *Optimality.* Finding $\mathbf{x}^*$, $Z(\mathbf{x}^*) = \text{MIN } Z(\mathbf{x})$, is the logical conclusion of the iterative search sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}^*$. However, in moving beyond just discovering an optimal solution to the next step of deriving a response surface approximation and characterizing the region of optimality, $\mathbf{x}^*$ plays a crucial role as the experimental design *centerpoint*. Also, in the case of multiple optimal solutions, the range of the elements of n optimal vectors $\mathbf{x}^*_1, \mathbf{x}^*_2, \ldots, \mathbf{x}^*_n$ provides guidance on how to scale the elements of the independent vector $\mathbf{x}$ in the formal design.

2. *Factor Screening.* As a byproduct of acquiring $\mathbf{x}^*$, the sequence of first-stage vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}^*$, and associated estimated responses $Z(\mathbf{x}_1), Z(\mathbf{x}_2), \ldots, Z(\mathbf{x}^*)$, form a data set for the purpose of *screening* the elements of $\mathbf{x}$ for

significant effects on the estimated response $Z(\mathbf{x})$. As Section 4.3 explains, such preliminary factor screening can help reduce the size of the formal experimental design used to calculate the response approximation.

The necessity of screening the components of $\mathbf{x}$ motivates the use of gradient-based search techniques over LP-based decomposition methods. Indeed, what the literature often views as disadvantages of such line-search techniques — slow convergence, sampling and comparison requirements — can prove advantageous (within reason) by providing ample experimental data to effectively reduce the number of independent variables.

As reviewed in Section 2.3.4 most search techniques find a directional vector $\mathbf{d}_k$ (not to be confused with the recourse cost vector $\mathbf{d}$ in (3.1b)) from an incumbent feasible solution $\mathbf{x}_k$ such that the following conditions hold for a minimization objective of (3.1):

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \rho_k \mathbf{d}_k, \tag{3.2a}$$

$$\rho_k > 0 \tag{3.2b}$$

$$\mathbf{A}\mathbf{x}_{k+1} = \mathbf{b},\ \mathbf{x}_{k+1} \geq \mathbf{0} \tag{3.2c}$$

$$Z(\mathbf{x}_{k+1}) \leq Z(\mathbf{x}_k). \tag{3.2d}$$

Note that (3.2) is equivalent to the projection operator often seen in the stochastic programming literature (see Section 2.3.4). Furthermore, given the known convex nature of $Z(\mathbf{x})$, (3.2) implies that for any descending directional vector $\mathbf{d}_k$ of any $\mathbf{x}_k$ where $Z(\mathbf{x}_k) > Z(\mathbf{x}^*)$, there exists an optimal ρ^*_k such that for any $\rho_j \neq \rho^*_k$, $Z(\mathbf{x}_k + \rho^*_k \mathbf{d}_k) < Z(\mathbf{x}_k + \rho_j \mathbf{d}_k)$. Therefore, the three search methods explored in

this research — *Geometric Simplex*, *Projected Gradient*, and *PARallel TANgents (PARTAN)* — must estimate two components: (1) the steepest directional descent vector $\mathbf{d}_k$ and (2) the optimal scalar multiple ρ^*_k.

3.2.2 Geometric Simplex

3.2.2.1 Introduction and Definitions

The geometric simplex search this effort implements follows the simplex search algorithm originally proposed by Spendley, Hext, and Himsworth (1962), as modified by Nelder and Mead (1965) and Barton and Ivey (1991). (For the remainder of this dissertation, the term *simplex* refers to the geometric simplex search as described in this section, and should not be confused with the linear programming optimization algorithm of the same name.) As explained by Barton and Ivey

> For a function of n parameters, the algorithm maintains a set of n+1 points in parameter space. This set of points defines a simplex in n dimensions. In two dimensions, the simplex would be a triangle; in three, a tetrahedron. The Spendley *et. al.* algorithm incorporates a regular simplex (i.e., all sides have the same length) which does not vary in size. The function is evaluated at each point of the simplex. The simplex then moves toward the optimum by reflecting the point with the worst function value through the centroid (average) of the remaining n points. In two dimensions, this can be visualized as flipping over a triangle to move it down a hill (Barton and Ivey 1991, 945).

Barton and Ivey show how Nelder-Mead modifies the Spendley *et al.* approach by allowing the shape of the simplex to change (thus allowing for quicker convergence). In terms of (3.2), this means $\mathbf{x}_k$ represents the vertex with the highest objective value in the current simplex, and its reflection through the

centroid of the remaining points determines the directional vector of descent $\mathbf{d}_k$. The simplex converges towards $\mathbf{x}^*$ by iteratively replacing $\mathbf{x}_k$ with a new solution $\mathbf{x}_{k+1}$ found along the directional vector $\mathbf{d}_k$ whenever $Z(\mathbf{x}_{k+1}) \leq Z(\mathbf{x}_k)$, and stops when meeting a predetermined termination criteria. Referring to Figure 3.2 below, these simplex moves include:

1. *Reflection*. The reflection vector extends beyond the centroid in the direction $\mathbf{d}_k$ to a candidate point $\mathbf{x}_r$ outside the boundaries of the current simplex.

2. *Expansion*. If $Z(\mathbf{x}_r) \leq Z(\mathbf{x}_k)$, the algorithm finds another candidate point $\mathbf{x}_e$ further away from the simplex in the direction $\mathbf{d}_k$, and compares $Z(\mathbf{x}_r)$ to $Z(\mathbf{x}_e)$.

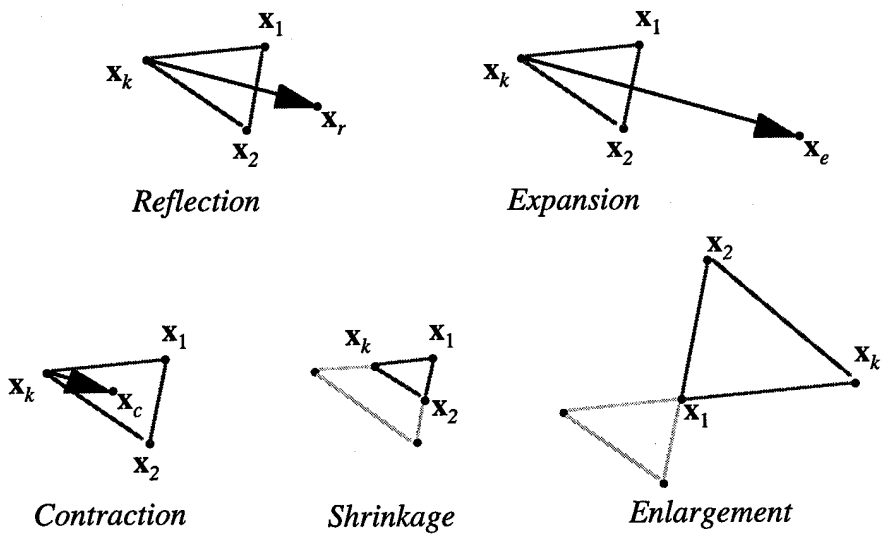


Fig. 3.2. Geometric Simplex Moves

3. *Contraction.* If the reflection point $Z(\mathbf{x}_r) > Z(\mathbf{x}_k)$, the algorithm finds another candidate point $\mathbf{x}_c$ (using $\mathbf{d}_k$) either closer to the centroid, or within the simplex boundaries, and compares $Z(\mathbf{x}_c)$ to $Z(\mathbf{x}_k)$.

4. *Shrinkage.* If both the reflection point $Z(\mathbf{x}_r) > Z(\mathbf{x}_k)$ and the contraction point $Z(\mathbf{x}_c) > Z(\mathbf{x}_k)$, then the algorithm shrinks the simplex by moving its points closer to the best vertex, while keeping the best vertex constant (Barton and Ivey 1991).

Since preliminary research indicates that item (4) often causes a premature collapse of the simplex when used on (3.1a) this study adds a fifth option:

5. *Enlargement.* Expand the simplex by extending its points through the best vertex to a greater distance on the other side (while keeping the best vertex constant). In effect, flipping and expanding the simplex about its best point.

Item (5) also has the added benefit of providing additional sampling of the parameter space needed for preliminary factor screening. Finally, as explained shortly the simplex search this dissertation implements restricts itself to two entering candidate evaluations: an interior point $\mathbf{x}_c$ halfway between the leaving candidate $\mathbf{x}_k$ and the centroid; and, an exterior point $\mathbf{x}_e$ outside the simplex half-way to a boundary constraint defined by $\boldsymbol{X}$.

> **Definition.** Let a *point* or *vertex* in the simplex be a feasible vector $\mathbf{x}$ of full dimension $n_{(3.1a)}$ and $Z(\mathbf{x})$ the objective function value as defined in

(3.1). Defining $I = n_{(3.1a)} + 1$ to be the number of vertices in the simplex, let i index $\mathbf{x}_i$ such that $Z(\mathbf{x}_1) \leq Z(\mathbf{x}_2) \leq \ldots \leq Z(\mathbf{x}_i) \leq \ldots \leq Z(\mathbf{x}_I)$.

Definition. Let $X_k = \{\mathbf{x}_i: \mathbf{x}_i \in \boldsymbol{X}, i = 1, \ldots, I\}$ represent the k^{th} simplex for (3.1), and by previous definition $X_k \subseteq \boldsymbol{X}$. Furthermore, letting $\mathbf{x}_{i,k}$ represent $\mathbf{x}_i \in X_k$, then for any entering candidate $\mathbf{x}_{j,k+1} \notin X_k$ and leaving candidate $\mathbf{x}_{i,k} \notin X_{k+1}$, $Z(\mathbf{x}_{j,k+1}) \leq Z(\mathbf{x}_{i-1,k})$.

The last definition states that the objective function value for the entering vertex must be better than the *next* best vertex value from the leaving candidate in the current simplex. This comparison helps prevent the search from stalling, as well as avoiding using the best vertex ($\mathbf{x}_1$) as a leaving candidate.

The simplex search algorithm can now be stated using these definitions:

Definition. Let $\mathbf{c}_{i,k}$ represent the *centroid* when $\mathbf{x}_{i,k}$ is the leaving vertex in simplex X_k. Define $c^j_{i,k} = (I - 1)^{-1} (x^j_{1,k} + x^j_{2,k} + \ldots + x^j_{i-1,k} + x^j_{i+1,k} + \ldots + x^j_{I,k})$ for $j = 1, \ldots, n_{(3.1a)}$. Since $\mathbf{c}_{i,k}$ is a convex combination of simplex vertices, $\mathbf{c}_{i,k} \in \boldsymbol{X}$.

Definition. Let $\mathbf{r}_{i,k}$ represent the *reflection vector* when $\mathbf{x}_{i,k}$ is the leaving vertex in simplex X_k and $\mathbf{c}_{i,k}$ is its associated centroid. Define $\mathbf{r}_{i,k} = \mathbf{c}_{i,k} - \mathbf{x}_{i,k}$. Since $\mathbf{c}_{i,k} \in \boldsymbol{X}$ and $\mathbf{x}_{i,k} \in \boldsymbol{X}$, $\mathbf{A}\mathbf{r}_{i,k} = \mathbf{A}\mathbf{c}_{i,k} - \mathbf{A}\mathbf{x}_{i,k} = \mathbf{b} - \mathbf{b} = \mathbf{0}$.

Definition. Let ρ_k represent a scalar multiple of the reflection vector $\mathbf{r}_{i,k}$ such that the entering vector $\mathbf{x}_{j,k+1} = \mathbf{c}_{i,k} + \rho_k \mathbf{r}_{i,k}$, and $\rho_k \geq 0$.

Definition. Let $R_k = \text{MIN}\left\{ \left| \frac{x^j_{ik}}{r^j_{ik}} \right| \text{ for } j = 1 \ldots n_{(3.1a)} \text{ and } r^j_{ik} < 0 \right\}$. Since $\mathbf{A}\mathbf{r}_{i,k} = \mathbf{0}$, the entering vector $\mathbf{x}_{j,k+1}$ will be feasible to $\mathbf{A}\mathbf{x} = \mathbf{b}$ *for any value* of ρ_k. However, since $\mathbf{x} \geq \mathbf{0}$ and $\rho_k \geq 0$, the only way any element of $\mathbf{x}_{j,k+1}$ can be less than zero is if $r^j_{i,k} < 0$. Therefore, R_k represents the maximum scalar multiple of $\mathbf{r}_{i,k}$ such that $\mathbf{x}_{j,k+1} \in X$, i.e., $\rho_k \leq R_k$.

3.2.2.2 Geometric Simplex Algorithm

STEP 0. *(Initialization).* Establish the first simplex vertex using the optimal solution $\mathbf{x}_{ev}$ from the expected value approximation. Randomly select remaining I - 1 vertices labeled $\mathbf{x}_0$ for initial simplex X_1. Estimate $Z(\mathbf{x}_i)$ and re-index on i using the relationship $Z(\mathbf{x}_1) \leq Z(\mathbf{x}_2) \leq \ldots \leq Z(\mathbf{x}_i) \leq \ldots \leq Z(\mathbf{x}_I)$. Set simplex counter index $k = 1$ and $i = I$. Set N (the terminating number of vertices) and vertex counter $n = 1$.

STEP 1. If $n > N$ STOP. Otherwise, find $\mathbf{c}_{i,k}$ and $\mathbf{r}_{i,k}$. Calculate two entering candidates: $\mathbf{x}_c = \mathbf{c}_{i,k} - .5\mathbf{r}^s_{i,k}$ and $\mathbf{x}_e = \mathbf{c}_{i,k} + .5R_k\mathbf{r}^s_{i,k}$. Compare $Z(\mathbf{x}_{i-1,k})$ to MIN $\{Z(\mathbf{x}_e), Z(\mathbf{x}_c)\}$ using the following guidelines.

If MIN$\{Z(\mathbf{x}_c), Z(\mathbf{x}_e)\} \leq Z(\mathbf{x}_{i-1,k})$ go to STEP 2.

If MIN$\{Z(\mathbf{x}_c), Z(\mathbf{x}_e)\} > Z(\mathbf{x}_{i-1,k})$ go to STEP 3.

STEP 2. Replace $Z(\mathbf{x}_{i,k})$ with MIN$\{Z(\mathbf{x}_c), Z(\mathbf{x}_e)\}$, and $\mathbf{x}_{i,k}$ with appropriate $\mathbf{x}_e$ or $\mathbf{x}_c$. Set $k = k + 1$. Re-index on i using the relationship $Z(\mathbf{x}_1) \leq Z(\mathbf{x}_2) \leq \ldots \leq Z(\mathbf{x}_i) \leq \ldots \leq Z(\mathbf{x}_I)$. Set $i = I$, and increment $n = n + 1$. Go to STEP 1.

STEP 3. Set $i = i - 1$. If $i = 1$ go to STEP 4; otherwise, go to STEP 1.

STEP 4. Based on the number of previous visits to this step do the following:

Shrinkage. Shrink the simplex by contracting $\mathbf{x}_2, \ldots, \mathbf{x}_I$ to one-half their current respective distance from $\mathbf{x}_1$ if *first* or *third* visit to STEP 4.

Enlargement. Enlarge the simplex by moving $\mathbf{x}_2, \ldots, \mathbf{x}_I$ *through* $\mathbf{x}_1$ to nine-tenths their current respective distance to the opposing feasible boundary if *second* visit to STEP 4.

Re-initialization. Otherwise re-initialize simplex by following the procedures in STEP 0 *except*: (1) Retain current $\mathbf{x}_1$ instead of $\mathbf{x}_{ev}$; and, (2) Retain current value for n. Go to STEP 1.

3.2.2.3 Implementation

The termination criteria of a preset number of vertices highlights the major drawback of the simplex search — *its inability to confirm an optimal solution.* Instead, it must *assume* an optimal solution has been found through other methods, such as an absence of any improving moves or reaching a predetermined standard error of the response estimate (Barton and Ivey 1991). In theory, the stopping criteria would concur with $Z(\mathbf{x}_{1,k}) = Z(\mathbf{x}^*)$; however, in practice the variation of the response estimator $\hat{Z}_s(\mathbf{x})$ can cause false convergence. Quoting Barton and Ivey again

> The Nelder-Mead algorithm is widely used for simulation optimization, where the functions it optimizes are often subject to random noise. The algorithm is robust to small inaccuracies or stochastic perturbations in function values. This is because the method uses only the ranks of the function values to determine the next move, not the function values

> themselves. Perturbations that do not change the ranks of the values will have no effect on the algorithm's search trajectory.
>
> If this noise is substantial, it will lead to inappropriate rescaling operations, resulting in false convergence. Empirically, this problem often manifests itself as inappropriate shrink steps. Once begun, this reduction in the simplex size can reduce the variance of the simplex function values below the system's inherent variability before the optimum region has been reached (Barton and Ivey 1991, 946-947).

Unfortunately, preliminary tests showed that the variance of $h(\mathbf{x},\omega,\mathbf{T})$ can indeed be substantial, and thus adversely affect the ordinal ranking of $\hat{Z}_s(\mathbf{x}_i)$ the simplex method depends upon to find $\mathbf{d}_k$ and ρ_k. However, these tests also indicate that the simplex search finds the region of optimality fairly quickly if, as in STEP 4, *it avoids a premature collapse through (1) periodic enlargement and (2) non-repetitive shrinkage.*

The termination criteria, and ultimately the simplex technique itself, assumes that the scope of the search will provide enough information about the region of optimality to insure a positive definite fit for the final response surface approximation of $Z(\mathbf{x})$. This procedure also assumes that the subsequent investigation of the minimum ridge estimates (see Section 4.4) will further characterize the region of optimality, and identify any improvements over the incumbent solution. Finding the optimal solution in this manner follows the main idea of steepest descent methods in the response surface literature, except the simplex method has been substituted for the first-order designs (see Montgomery 1984, or Box and Draper 1987). Finally, variance reduction techniques explored in Section 4.2 may reduce the variation of $h(\mathbf{x},\omega,\mathbf{T})$ to the "...small inaccuracies or

stochastic perturbations in function values..." Barton and Ivey claim as acceptable for the simplex search.

3.2.2.4 Example Simplex Search for APL1P

Figure 3.3 illustrates the first two iterations of a 20-iteration simplex search path for the APL1P problem, and highlights the initial simplex shape and its form after the entry of $\mathbf{x}_2$ (Table 3.1 provides the simplex data for all 20 iterations). As a two-dimensional space, APL1P requires three vertices, with the initial simplex (outlined by the upper triangle) using the expected value approximation solution ($\mathbf{x}_{ev}$) and two randomly selected ones ($\mathbf{x}_0$). The first iteration replaces the most expensive $\mathbf{x}_0$ with $\mathbf{x}_1$ through an expansion move, while the second iteration candidate $\mathbf{x}_2$ replaces the other $\mathbf{x}_0$ with a contraction move. After two moves the simplex X_2 (outlined by the lower triangle) is significantly closer to the region of optimality. The simplex continues to contract until undergoing a shrink-enlarge-shrink cycle after iteration 17. After two additional contractions, the simplex re-initializes and undergoes one additional contraction before terminating. The final near-optimal solution $\mathbf{x}' = \mathbf{x}_{19} =$ (1708, 1685) and $Z(\mathbf{x}') = 24647.49$, compared to $\mathbf{x}^* =$ (1800, 1570) and $Z(\mathbf{x}^*) =$ 24642.29.

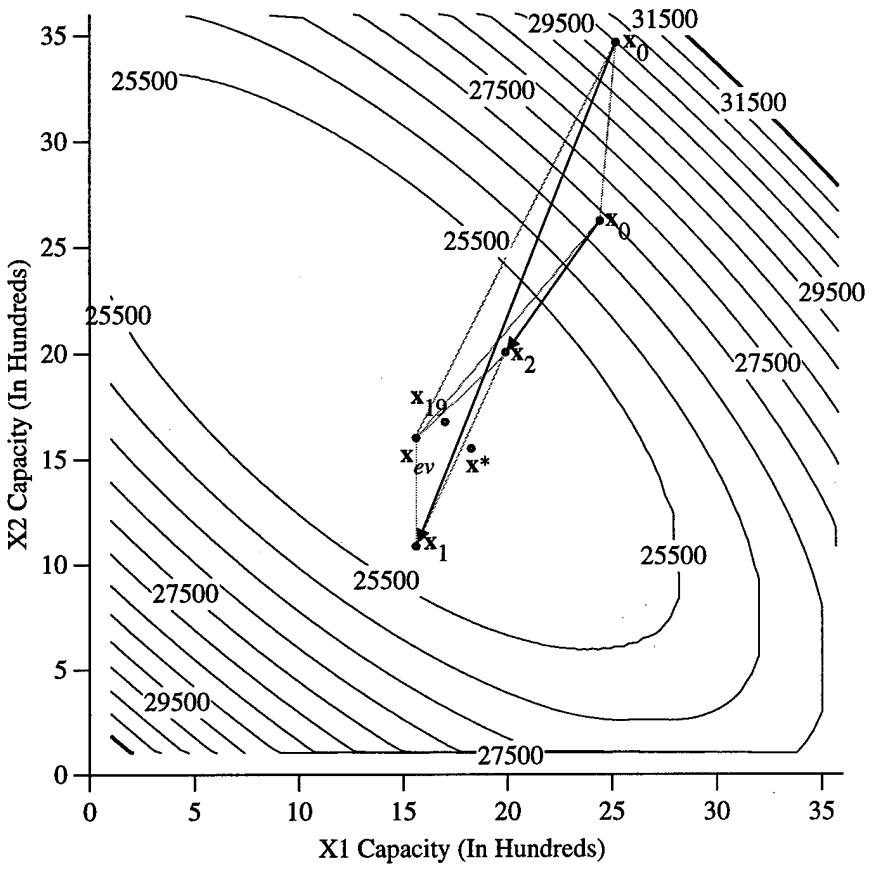


Figure 3.3. Example Geometric Simplex Moves for APL1P

3.2.3 Projected Gradient

3.2.3.1 Introduction and Definitions

The relationship described in (3.2) requires three basic operations at each iteration: (1) finding a directional vector $\mathbf{d}_k$ satisfying $\mathbf{A}\mathbf{d}_k = 0$; (2) deriving an optimal ρ^*_k such that for any $\rho_j \neq \rho^*_k$, $Z(\mathbf{x}_k + \rho^*_k\mathbf{d}_k) \leq Z(\mathbf{x}_k + \rho_j\mathbf{d}_k)$; and, (3)

TABLE 3.1
EXAMPLE GEOMETRIC SIMPLEX MOVES FOR APL1P
(RANDOM SEED = 440908571)

k	$\mathbf{x}$	$Z(\mathbf{x})$	Simplex Move	Replaces
0	1529, 1625	24698.47	Initial Vertex ($\mathbf{x}_{ev}$)	—
0	2437, 2680	26824.78	Initial Vertex (Random Pick)	—
0	2502, 3462	28756.62	Initial Vertex (Random Pick)	—
1	1557, 1077	25216.05	Expand $\mathbf{x}_e$	3
2	1990, 2016	24997.72	Contract $\mathbf{x}_c$	3
3	1658, 1449	24731.60	Contract $\mathbf{x}_c$	3
4	1792, 1776	24676.73	Contract $\mathbf{x}_c$	3
5	1659, 1575	24670.40	Contract $\mathbf{x}_c$	3
6	1623, 1650	24661.63	Contract $\mathbf{x}_c$	3
7	1718, 1694	24647.60	Contract $\mathbf{x}_c$	3
8	1666, 1624	24657.75	Contract $\mathbf{x}_c$	3
9	1660, 1655	24653.83	Contract $\mathbf{x}_c$	3
10	1677, 1649	24651.15	Contract $\mathbf{x}_c$	3
11	1679, 1663	24649.44	Contract $\mathbf{x}_c$	3
12	1688, 1664	24648.67	Contract $\mathbf{x}_c$	3
13	1691, 1672	24648.20	Contract $\mathbf{x}_c$	3
14	1696, 1673	24647.70	Contract $\mathbf{x}_c$	3
15	1699, 1678	24647.69	Contract $\mathbf{x}_c$	3
16	1702, 1680	24647.59	Contract $\mathbf{x}_c$	3
17	1709, 1689	24647.50	Contract $\mathbf{x}_c$	3
0	—	—	Shrink	—
0	—	—	Enlarge	—
0	—	—	Shrink	—
18	1707, 1685	24647.53	Contract $\mathbf{x}_c$	3
19	1708, 1685	24647.49	Contract $\mathbf{x}_c$	3
0	1708, 1685	24647.49	Best Vertex from Prev. Simplex	—
0	1722, 2361	25134.60	New Vertex (Random Pick)	—
0	1176, 1015	25727.43	New Vertex (Random Pick)	—
20	1445, 1519	24792.78	Contract $\mathbf{x}_c$	3

recognizing when $Z(\mathbf{x}_k + \rho^*_k \mathbf{d}_k) = Z(\mathbf{x}^*) = \text{MIN } Z(\mathbf{x})$. The projected gradient method adopted by this study accomplishes these operations through a synthesis of the following ideas from the literature:

1. Using a concept first explored by Ermoliev (1983, 1988) within the context of the recourse problem, the dual variables provide an

unconstrained steepest descent gradient $\nabla Z(\mathbf{x}_k)$ for the first-stage variables $\mathbf{x}$. (Also see Murty 1983 for the theoretical LP background on dual variables as gradients.)

2. Gaivoronski's (1988) averaging, or statistical estimation, technique for cases where the estimate of the unconstrained steepest descent gradient is a function of a random variable; i.e., $\nabla Z(\mathbf{x},\omega,T)$.

3. Active set methods, whereby any 'true' equality constraints (i.e., no slack variables present), plus any inequality constraints whose slack variables currently equal 0, define the 'working surface' upon which a projected steepest descent gradient $\nabla Z(\mathbf{x}_k)$ produces a direction vector $\mathbf{d}_k$ such that $\mathbf{A}\mathbf{d}_k = \mathbf{0}$.

4. Estimating the stepsize variable ρ_k by fitting a quadratic model along the direction of descent defined by $\mathbf{d}_k$ using standard linear regression (Fu 1994, Luenberger 1989).

Applying steepest descent methods from the non-linear programming literature to the class of problems (3.1) describes comes with several caveats. First, in a deterministic setting active set methods can determine whether $\mathbf{x}_k$ is optimal, or otherwise provide active constraint relaxation guidelines to continue the search for $\mathbf{x}^*$ under the Kuhn-Tucker Theorem (Luenberger 1989). By contrast, the *statistical estimation* (when necessary) of the unconstrained descent gradient $\nabla Z(\mathbf{x}_k,\omega,T)$ implies the presence of error similar to that associated with

$\hat{Z}_s(\mathbf{x})$. Second, the non-differentiable property of $E[h(\mathbf{x},\omega,\mathbf{T})]$ suggests that — barring multiple optimality — $\nabla Z(\mathbf{x}_k)$ will never equal zero. Consequently, terminating the projected gradient algorithm will require the heuristic stopping rules proposed shortly.

The following definitions explain the terminology of the PROJECTED GRADIENT ALGORITHM.

Definition. For the vector $\mathbf{x}_k \in E^{n(3.1a)}$ let $\mathbf{A}_k$ represent the matrix composed of the active rows from $\mathbf{Ax} = \mathbf{b}$, and rows from $\mathbf{x} \geq \mathbf{0}$ where $x^i_k = 0$, $i = 1, \ldots, n_{(3.1a)}$.

Definition. Let $\nabla z(\mathbf{x}_k,\omega_i,\mathbf{T}_i)$ be the unconstrained gradient for the i^{th} realization of $\mathbf{x}_k$.

Definition. Let $\nabla\hat{Z}_s(\mathbf{x}_k)$ represent the unbiased estimate of $\nabla Z(\mathbf{x}_k)$ for sampling technique s.

Definition. Let $\mathbf{d}_k$ represent a feasible direction of improvement; i.e. $\nabla\hat{Z}_s(\mathbf{x}_k)\mathbf{d}_k < 0$ and $\mathbf{Ad}_k = \mathbf{0}$.

Definition. Let $\mathbf{J}_k$ be the projection matrix where $\mathbf{d}_k = -\mathbf{J}_k \cdot \nabla\hat{Z}_s(\mathbf{x}_k)$.

Definition. Let P_k represent the minimum scalar value of $\mathbf{d}_k$ at which a non-active non-negativity constraint $x^j > 0$, $x^j \notin \mathbf{A}_k$, goes to 0; i.e., the vector $\mathbf{x}_{k+1} = \mathbf{x}_k + \rho_k\mathbf{d}_k$ such that $\mathbf{x}_{k+1} \in \boldsymbol{X}$ only when $0 \leq \rho_k \leq P_k$.

Definition. Let $\gamma \in E^{n(3.1a)}$ be the stopping criteria for projected gradient search.

Definition. Let Q represent the number of independent estimates of $Z(\mathbf{x})$ taken at equal distances along the vector starting at $\mathbf{x}_k$ in the direction of $\mathbf{d}_k$ and ending at $\mathbf{x}_k + P_k\mathbf{d}_k$. Then where $q_i = (Q - 1)^{-1} \cdot (i - 1)$, $i = 1, \ldots, Q$, let $\mathbf{x}_{k,qi} = \mathbf{x}_k + q_iP_k\mathbf{d}_k$, and $Z(\mathbf{x}_{k,qi}) = Z(\mathbf{x}_k + q_iP_k\mathbf{d}_k)$.

Definition. Let $\hat{Z}_{LN}(\mathbf{x}_k,q) = Z(\mathbf{x}_k + qP_k\mathbf{d}_k) + \varepsilon_{k,LN}$ and $\hat{Z}_{QD}(\mathbf{x}_k,q) = Z(\mathbf{x}_k + qP_k\mathbf{d}_k.) + \varepsilon_{k,QD}$ represent first- and second-order polynomial approximations, respectively, of $Z(\mathbf{x}_k + q_iP_k\mathbf{d}_k)$ as functions of $\mathbf{x}_k$ and q, $0 \leq q \leq 1$.

3.2.3.2 Projected Gradient Algorithm

STEP 0. *(Initialization)* Set $k = 1$ and $n = 1$. Select $\mathbf{x}_k$. Select stopping criteria γ and maximum number of iterations N. Construct active set matrix $\mathbf{A}_k$ using row vectors from $\mathbf{Ax} = \mathbf{b}$, and from $\mathbf{x} \geq \mathbf{0}$ for any $x^i_k = 0$. Estimate and record $Z(\mathbf{x}_k)$.

STEP 1. Calculate $\nabla\hat{Z}_s(\mathbf{x}_k)$. Find $\mathbf{d}_k$ by projecting $\nabla\hat{Z}_s(\mathbf{x}_k)$ onto $\mathbf{A}_k$ using projection matrix $\mathbf{J}_k$. If $\mathbf{d}_k \leq |\gamma|$ or $n > N$, STOP. Otherwise, find P_k, select Q, estimate and record $Z(\mathbf{x}_{k,qi})$, $i = 1, \ldots, Q$.

STEP 2. Derive $\hat{Z}_{LN}(\mathbf{x}_k,q)$ and $\hat{Z}_{QD}(\mathbf{x}_k,q)$. Select q^* under the following guidelines:

Quadratic Significance. If $\hat{Z}_{QD}(\mathbf{x}_k,q)$ significant find q^* by setting derivative of $\hat{Z}_{QD}(\mathbf{x}_k,q)$ with respect to q equal to zero and solve. If $q^* < 0$ set $q^* = .01$. If $q^* > 1$, set $q^* = 1.0$.

Linear Significance. If $\hat{Z}_{LN}(\mathbf{x}_k,q)$ has a negative slope set $q^* = 1.0$; otherwise set $q^* = .01$.

Neither Fit Significant. Set $q^* = .01$

STEP 3. Set $\mathbf{x}_{k+1} = \mathbf{x}_k + q^*\mathrm{P}_k\mathbf{d}_k$, $\mathbf{A}_{k+1} = \mathbf{A}_k$, and $\mathbf{J}_{k+1} = \mathbf{J}_k$. Set $k = k + 1$. Estimate $Z(\mathbf{x}_k)$ and $\mathbf{x}_k$. Check $\mathbf{x}_k$ for either (1) $x^i_k > 0$ where $x^i_{k-1} = 0$ or (2) $x^i_k = 0$ where $x^i_{k-1} > 0$. If (1) occurs remove x^i_k from $\mathbf{A}_k$; if (2) occurs add x^i_k to $\mathbf{A}_k$. If either (1) or (2) occur (i.e., $\mathbf{A}_k \neq \mathbf{A}_{k-1}$), recalculate $\mathbf{P}_k$. Return to STEP 1.

3.2.3.3 Theoretical Development and Implementation

The structure of (3.1) requires that the respective dual variable information for the **x** variables must come from the recourse problem (3.1b). Given the following primal (3.3a) and dual (3.3b) formulations

$$\begin{aligned} h(\mathbf{x},\omega,\mathbf{T}) \;=\; & \text{MIN } \mathbf{dy} \\ & \text{s.t.} \quad \mathbf{Tx} + \mathbf{Wy} = \omega \\ & \qquad\qquad \mathbf{y} \geq \mathbf{0} \end{aligned} \tag{3.3a}$$

$$\begin{aligned} h(\mathbf{x},\omega,\mathbf{T}) \;=\; & \text{MAX } \pi(\omega - \mathbf{Tx}) \\ & \text{s.t.} \quad \pi\mathbf{W} \leq \mathbf{d} \\ & \qquad\qquad \pi \text{ Unrestricted} \end{aligned} \tag{3.3b}$$

Ermoliev (1983, 1988) shows that for the i^{th} realization of ω and $\mathbf{T}$

$$-\nabla z_{ik} = \mathbf{c} + \pi_{ik}\mathbf{T} \tag{3.4}$$

and for $\mathbf{x}_k$ the unbiased estimate of the unconstrained steepest descent gradient for sample size N is

$$-\nabla \hat{Z}_s(\mathbf{x}_k) = \frac{1}{N}\sum_{i=1}^{N} -\nabla z_{ik}. \tag{3.5}$$

Unfortunately, $\mathbf{A}[-\nabla \hat{Z}_s(\mathbf{x}_k)]$ may not equal $\mathbf{0}$; therefore, a directional vector $\mathbf{d}_k$ that remains feasible beyond just the tangential point of the constraints to $Z(\mathbf{x})$ must be found by projecting $-\nabla \hat{Z}_s(\mathbf{x}_k)$ onto the set of active constraints. This active set method of gradient projection follows the description in Luenberger (1989) as originally suggested by Rosen (1960) and Gill, Murray, and Wright (1981).

From the definition of a plane, the row $\mathbf{A}^{i\bullet}{}_k$ of $\mathbf{A}_k$ is a normal vector to the subspace defined by the constraint. Therefore, any vector in the Euclidean space $E^{n(3.1a)}$ (specifically $-\nabla \hat{Z}_s(\mathbf{x}_k)$) can be defined as a linear combination of $\mathbf{d}_k$ and the rows in $\mathbf{A}_k$. The projection matrix $\mathbf{J}$ can then be derived starting with

$$-\nabla \hat{Z}_s(\mathbf{x}_k) = \mathbf{d}_k + (\mathbf{A}_k)^{\mathrm{T}}\lambda_k. \tag{3.6}$$

Multiplying (3.6) by $\mathbf{A}_k$ and using $\mathbf{A}_k\mathbf{d}_k = 0$ gives

$$-\mathbf{A}_k\nabla \hat{Z}_s(\mathbf{x}_k) = \mathbf{A}_k\mathbf{d}_k + \mathbf{A}_k(\mathbf{A}_k)^{\mathrm{T}}\lambda_k = \mathbf{A}_k(\mathbf{A}_k)^{\mathrm{T}}\lambda_k \tag{3.7}$$

and multiplying (3.7) by $[\mathbf{A}_k)(\mathbf{A}_k)^{\mathrm{T}}]^{-1}$ reduces to

$$\lambda_k = [\mathbf{A}_k(\mathbf{A}_k)^{\mathrm{T}}]^{-1}\mathbf{A}_k\nabla \hat{Z}_s(\mathbf{x}_k). \tag{3.8}$$

Substituting λ_k from (3.8) into (3.6) gives

$$\mathbf{d}_k = -\mathbf{J}_k \nabla \hat{Z}_s(\mathbf{x}_k) \tag{3.9}$$

where the projection matrix $\mathbf{J}_k$ is defined as

$$\mathbf{I} - (\mathbf{A}_k)^T[\mathbf{A}_k(\mathbf{A}_k)^T]^{-1}\mathbf{A}_k. \tag{3.10}$$

Accordingly, for each $\mathbf{x}_k$ a directional vector $\mathbf{d}_k$ (where $\mathbf{A}\mathbf{d}_k = \mathbf{0}$) can be found using the unconstrained gradient estimate $-\nabla \hat{Z}_s(\mathbf{x}_k)$, and the projection matrix $\mathbf{J}_k$ (3.10) based on the current active matrix $\mathbf{A}_k$, with the relationship (3.9) (Luenberger 1989).

The directional vector $\mathbf{d}_k$ guarantees that the constraints in the *current active set* $\mathbf{A}_k$ will not be violated. Specifically, for *any* $\rho_k \geq 0$ the vector $\mathbf{x}_{k+1} = \mathbf{x}_k + \rho_k \mathbf{d}_k$ will be feasible to $\mathbf{A}\mathbf{x} = \mathbf{b}$. However, following an identical argument from the GEOMETRIC SIMPLEX ALGORITHM, since $\mathbf{x} \geq \mathbf{0}$ and $\rho^*_k \geq 0$ the only way any x^i_{k+1} can be infeasible is if $d^i_k < 0$ (since $x^i_{k+1} = x^i_k + \rho_k d^i_k < 0$). Therefore, it follows directly that

$$P_k = \text{MIN}\left\{ \left| \frac{x^i_k}{d^i_k} \right| \text{ for } i = 1 \ldots n_{(3.1a)} \text{ and } d^i_k < 0 \right\}. \tag{3.11}$$

Furthermore, given the known convexity of $Z(\mathbf{x})$, there exists an optimal multiplier $0 \leq \rho^*_k \leq P_k$ for $\mathbf{d}_k$ such that for any $0 \leq \rho_j \leq P_k$, $\rho_j \neq \rho^*_k$, $Z(\mathbf{x}_k + \rho^*_k \mathbf{d}_k) < Z(\mathbf{x}_k + \rho_j \mathbf{d}_k)$. The literature refers to finding this optimal ρ^*_k as the stepsize problem, which the following section addresses.

Fu (1994), in the context of response surface sequential search (also see Box and Draper 1987), suggests solving the step-size problem by formulating a line search defined by the descent gradient as a one-dimensional optimization problem and fitting a second-order polynomial. This approach offers a special appeal for finding MIN $Z(\mathbf{x})$ due to $Z(\mathbf{x})$'s known convexity and global optimal characteristics; and, partially addresses concerns expressed in the literature over selecting inefficiently small or large step-sizes (Sivazlian and Stanfel 1975, Luenberger 1989). Therefore, this research extends Fu's idea by testing for quadratic significance versus a first-order fit and selecting the best step-size within the constraints defined by $\boldsymbol{X}$.

The basic idea involves deriving a linear and quadratic approximation of $Z(\mathbf{x}_k)$ as a function of q; i.e.,

$$\hat{Z}_{LN}(\mathbf{x}_k,q) = \beta_0 + \beta_1 q + \varepsilon_{k,LN} \tag{3.12}$$

and

$$\hat{Z}_{QD}(\mathbf{x}_k,q) = \beta_0 + \beta_1 q + \beta_2 q^2 + \varepsilon_{k,QD} \tag{3.13}$$

where $0 \leq q \leq 1$, based on data derived from an equidistant sampling of the directional vector $\mathbf{d}_k$ from the incumbent solution $\mathbf{x}_k$ to the bounds defined by $\mathbf{x} \geq \mathbf{0}$. For example, if $Q = 6$, then the line segment starting at $\mathbf{x}_k$ and ending at $\mathbf{x}_k + P_k\mathbf{d}_k$ will be sampled at intervals of 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0 multiples of P_k, and (3.12) and (3.13) derived using q_i as the independent variable and $Z(\mathbf{x}_{k,qi})$ as the dependent variable.

The algorithm allows for several subjective interpretations. First, the selection of Q in STEP 1 involves a trade-off between more accurate regression estimates on the one hand, and additional computational time on the other. This research initially uses a standard factor of $Q = 6$ as a compromise, although the program code allows for $Q \leq 10$. Second, as in most regression studies the term 'significant' depends on the views of the analyst and the context of the analysis (see Draper and Smith 1981). Preliminary research indicates that the combined effects of the variance of $h(\mathbf{x},\omega,\mathbf{T})$ and the 'flatness' of the region of optimality can create lack of fit results near constraint boundaries or the optimal point. Consequently, requiring too high a fit would significantly lengthen the search process. Therefore, 'significance' in this study constitutes an R-Square fit greater than 0.9. Third, STEP 3 tracks the changes in $\mathbf{x}_k$ to avoid recalculating the projection matrix when there is no need to do so. Fourth, depending on its location in X the algorithm's quadratic approximation in STEP 3 can easily decide on either a negative value of q^*, or one greater than 1.0 — either case an obviously false estimate assuming the correct directional gradient. The former can occur as the search approaches $\mathbf{x}^*$, whereas the latter typically would happen whenever $Z(\mathbf{x}_k) >> Z(\mathbf{x}^*)$. These conditions would thus suggest using the predetermined increments 0.01 and 1.0, respectively.

Finally, a word about the termination criteria. Recognizing the limitations of estimating both $Z(\mathbf{x})$ and $\nabla Z(\mathbf{x}_k)$ by using pre-determined stopping criteria γ and N does not detract from stopping at a near-optimal solution $\mathbf{x}'$. Recalling the principal objective of deriving a response surface approximation of $Z(\mathbf{x})$ through experimental design, $Z(\mathbf{x}')$ or $Z(\mathbf{x}^*)$ represents *the starting point of the process.*

Ermoliev (1988) emphasizes this practical viewpoint from the context of finding the *optimal solution* using stochastic gradient methods; thus, it follows that by conducting this search within the more practical context of response surface approximation, terminating the search with a near-optimal solution is justified.

3.2.3.4 Example Projected Gradient for APL1P

Figure 3.4 (based on Table 3.2 below) provides a graphic illustration of the PROJECTED GRADIENT ALGORITHM applied to the APL1P problem. Starting with an initial solution $\mathbf{x}_1$ = (3600,3600), the algorithm very quickly moves to the region of optimality (as compared with the GEOMETRIC SIMPLEX ALGORITHM), but slows appreciably afterwards in finding the optimal point. This slowdown most likely results from less accurate estimates of q^* provided by the quadratic estimates of a relatively flat region in the direction of least sensitivity. Such inaccuracy manifests itself by the small *negative* estimates of q^* for $\mathbf{x}_4$ - $\mathbf{x}_{10}$; in

TABLE 3.2
EXAMPLE PROJECTED GRADIENT ITERATIONS FOR APL1P
(SCENARIOS = 1280, Q = 6)

k	$\mathbf{x}^\dagger$	π	Est. q^*	Act. q	R^2	$Z(\mathbf{x})^\dagger$
1	3600, 3600	3.28, 2.42	.613	.613	.990	32224.4
2	1394, 1969	-.094, .105	.079	.079	.995	24693.1
3	1532, 1815	-.103, .052	.044	.044	.999	24667.7
4	1623, 1769	-.048, .104	-.158	.010*	.995	24657.1
5	1632, 1751	-.050, .100	-.168	.010*	.995	24654.9
6	1640, 1733	-.050, .097	-.185	.010*	.995	24652.7
7	1649, 1716	-.081, .069	-.022	.010*	.999	24650.7
8	1669, 1700	-.068, -.007	-.183	.010*	.998	24649.3
9	1688, 1701	-.035, .022	-.019	.010*	.999	24648.6
10	1707, 1689	-.035, .022	—	—	—	24647.6

* - Manual Input † - $Z(\mathbf{x}^*)$ = 24642.3, $\mathbf{x}^*$ = (1800, 1570)

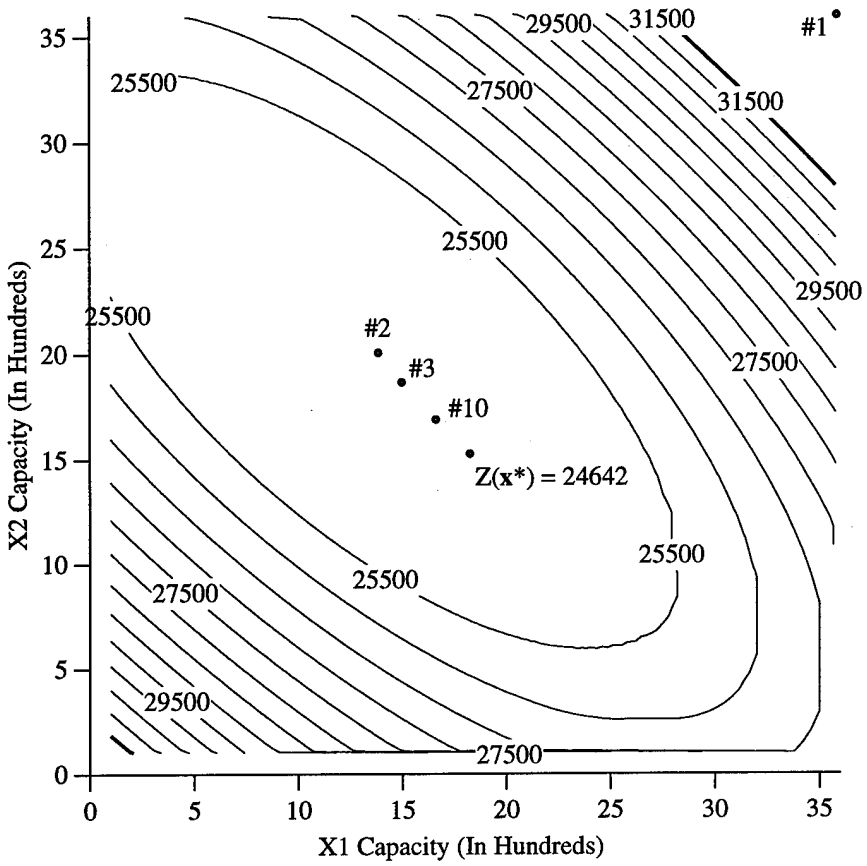


Figure 3.4. Example Projected Gradient Iterations for APL1P ($\mathbf{x}_4$ - $\mathbf{x}_9$ Omitted)

these cases the estimated quadratic fit places the minimum point *behind* the direction of descent (and thus requiring a .01 input for the stepsize). This phenomenon most likely occurs due to the incumbent $\mathbf{x}_k$ being so close to the optimal solution that most (or all) samplings of the directional vector $\mathbf{d}_k$ are greater than $Z(\mathbf{x}_k)$, thus forcing a least-squares estimate of a curve with minimal sampling in the opposite direction. Nonetheless, the algorithm provides accurate

directional information even as the stepsize problem prevents a quicker convergence.

3.2.4 Parallel Tangents (PARTAN)

3.2.4.1 Introduction and Definitions

Both Luenberger (1989) and Sivazlian and Stanfel (1975) discuss the special PARTAN procedure adapted by this dissertation. As shown in Figure 3.5, the basic idea involves finding a point $\mathbf{p}_k$ in the $\mathbf{x}$ parameter space using a steepest descent technique (in this case the PROJECTED GRADIENT ALGORITHM) from $\mathbf{x}_k$ such that $Z(\mathbf{p}_k) \leq Z(\mathbf{x}_k)$. Then, $\mathbf{x}_{k+1}$ is found by minimizing $Z(\mathbf{x})$ along the line defined by $\mathbf{x}_{k-1}$ and $\mathbf{p}_k$. Furthermore, at each iteration $\mathbf{x}_k$ is checked for optimality

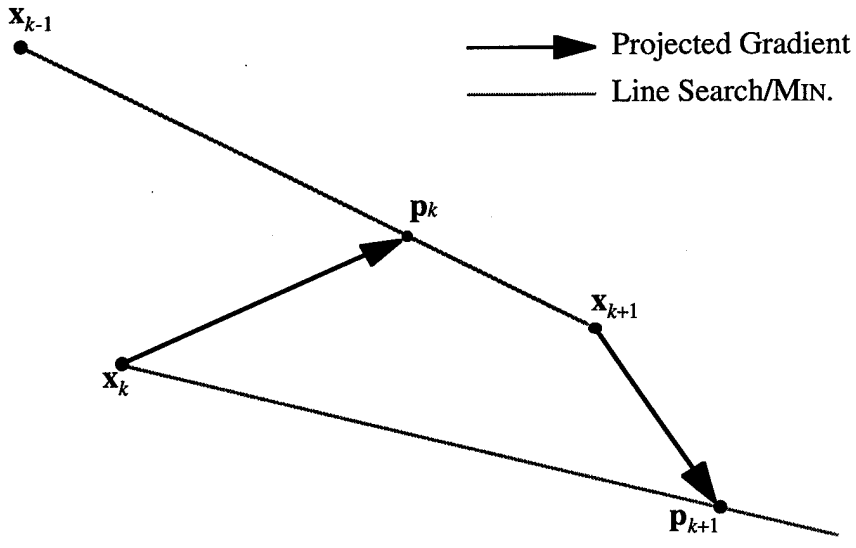


Figure 3.5. PARTAN Iteration

as it would be under the PROJECTED GRADIENT ALGORITHM, including any adjustment to its active matrix constraint $\mathbf{A}_k$. The formal definitions and algorithm are described below.

Definition. Let $\mathbf{p}_k$ be the next point in a PROJECTED GRADIENT ALGORITHM with respect to $\mathbf{x}_k$; i.e., $\mathbf{p}_k = \mathbf{x}_k + q^*_k \mathrm{P}_k \mathbf{d}_k$.

Definition. Let $\mathbf{t}_k$ represent the vector defined by $\mathbf{p}_k$ and $\mathbf{x}_{k-1}$; i.e., $\mathbf{t}_k = \mathbf{p}_k - \mathbf{x}_{k-1}$.

Definition. Let $\phi_k = \mathrm{MIN}\left\{ \left| \frac{\mathrm{x}^i_k}{\mathrm{t}^i_k} \right| \text{for } i = 1 \ldots n_{(3.1a)} \text{ and } \mathrm{d}^i_k < 0 \right\}$.

Definition. Let R represent the number of independent estimates of $Z(\mathbf{x})$ taken along the vector starting at $\mathbf{x}_{k-1}$ in the direction of $\mathbf{t}_k$ and ending at $\mathbf{p}_k + \phi_k \mathbf{t}_k$. Then where $r_i = (R - 1)^{-1} \cdot (i - 1)$, $i = 1, \ldots, R$, let $\mathbf{x}_{k+1,ri} = \mathbf{x}_{k-1} + r_i \phi_k \mathbf{t}_k$.

Definition. Let $\mathbf{x}_k$ be the current minimum point found by minimizing along the line defined by $\mathbf{x}_{k-2}$ and $\mathbf{p}_{k-1}$; i.e., $\mathbf{x}_k = \mathbf{x}_{k-2} + r^*_i \phi_{k-1} \mathbf{t}_{k-1}$.

3.2.4.2 PARTAN Algorithm

STEP 0. *(Initialization)* Select $\mathbf{x}_0 \in \boldsymbol{X}$. Estimate $\mathbf{x}_1 = \mathbf{x}_0 + q^*_0 \mathrm{P}_0 \mathbf{d}_0$ using the PROJECTED GRADIENT ALGORITHM. Set $k = 1$, $n = 1$, the maximum number of iterations N, and stopping criteria γ.

STEP 1. Calculate $\mathbf{p}_k = \mathbf{x}_k + q^*_k \mathrm{P}_k \mathbf{d}_k$ using the PROJECTED GRADIENT ALGORITHM. If $\mathbf{d}_k \leq |\gamma|$ or $n > N$ STOP. Otherwise find $\mathbf{x}_{k+1}$ by minimizing

along the *entire feasible line* defined by $\mathbf{x}_{k-1}$ and $\mathbf{p}_k$; i.e., find r^*_i using the quadratic fit procedure described in STEP 2 of the PROJECTED GRADIENT ALGORITHM such that $\mathbf{x}_{k+1} = \mathbf{x}_{k-1} + r^*_i \phi_k \mathbf{t}_k$. Set $k = k + 1$, $n = n + 1$, and repeat.

3.2.4.3 Theoretical Development and Implementation

One notable change from the PROJECTED GRADIENT ALGORITHM involves the line minimization requirement in STEP 1. In the PROJECTED GRADIENT ALGORITHM the quadratic fit involves the line defined along the descent gradient *forward* from the incumbent solution $\mathbf{x}_k$ to a constraint boundary. However, the PARTAN ALGORITHM pays the additional computational cost of searching the entire feasible line both *forward and behind* $\mathbf{p}_k$. This difference stems from the unique objectives and assumptions of the two algorithms. At each iteration the PROJECTED GRADIENT ALGORITHM focuses on the stepsize problem while assuming it has the correct directional descent vector; hence, the need for estimating a quadratic fit ahead of $\mathbf{x}_k$. By contrast, the PARTAN ALGORITHM's quadratic assumption avoids the stepsize problem through the use of parallel tangents; therefore, it concentrates instead on finding a good *global* quadratic estimate at $\mathbf{p}_k$.

The PARallel TANgent (PARTAN) approach extends steepest descent techniques to the special case where the non-linear function being estimated is a positive definite quadratic (Sivazlian and Stanfel 1975). Since the projected gradient method falls into the class of steepest descent algorithms, and given that

this research assumes Z(**x**) can be well approximated with a quadratic fit, PARTAN methods provide an obvious search technique to try. Indeed, as Luenberger shows, for a pure quadratic function PARTAN is equivalent to the conjugate gradient method, which itself has "...proved to be extremely effective in dealing with general objective functions and ...[is] considered among the best general purpose methods presently available (Luenberger 1989, 238)." Similarly Sivazlian and Stanfel (1975) show, in theory and for a pure n-dimensional quadratic, how PARTAN will find the optimal solution in just n-1 optimizations. They also note that for non-quadratic functions, PARTAN can still perform well in the reduced region of optimality where a quadratic approximation would be more accurate. Nonetheless, the single greatest theoretical advantage PARTAN brings to the problem (3.1a) is its strong global convergence properties. Quoting Luenberger again

> Each step of the process is at least as good as steepest descent; since going from $\mathbf{x}_k$ to ... [$\mathbf{p}_k$] ... is exactly steepest descent, and the additional move to $\mathbf{x}_{k+1}$ provides further decrease of the objective function. Thus global convergence is not tied to the fact that the process is restarted every n steps (Luenberger 1989, 256-257).

Although (3.1a) is neither deterministic or purely quadratic, the claims made on behalf of PARTAN justify investigating its performance within such a context.

3.2.4.4 Example PARTAN

Figure 3.6 shows the results of the PARTAN algorithm for the APL1P problem. Following Sivazlian and Stanfel's (1975) argument that where **x** has a dimension of $n_{(3.1a)}$ the global minimum should be found in $n_{(3.1a)}$ - 1 iterations of

the PARTAN algorithm, Figure 3.6 and its accompanying table show just one cycle. As expected, the PARTAN ALGORITHM exhibits a stronger convergence property than either the GEOMETRIC SIMPLEX or PROJECTED GRADIENT methods; it achieves in two steps what took the GEOMETRIC SIMPLEX ALGORITHM seven and the PROJECTED GRADIENT ALGORITHM five. Furthermore, the lines in Figure 3.6, coupled with the fact that $Z(\mathbf{x}_{k+1}) \approx Z(\mathbf{p}_k)$ and the absence of a pure quadratic fit,

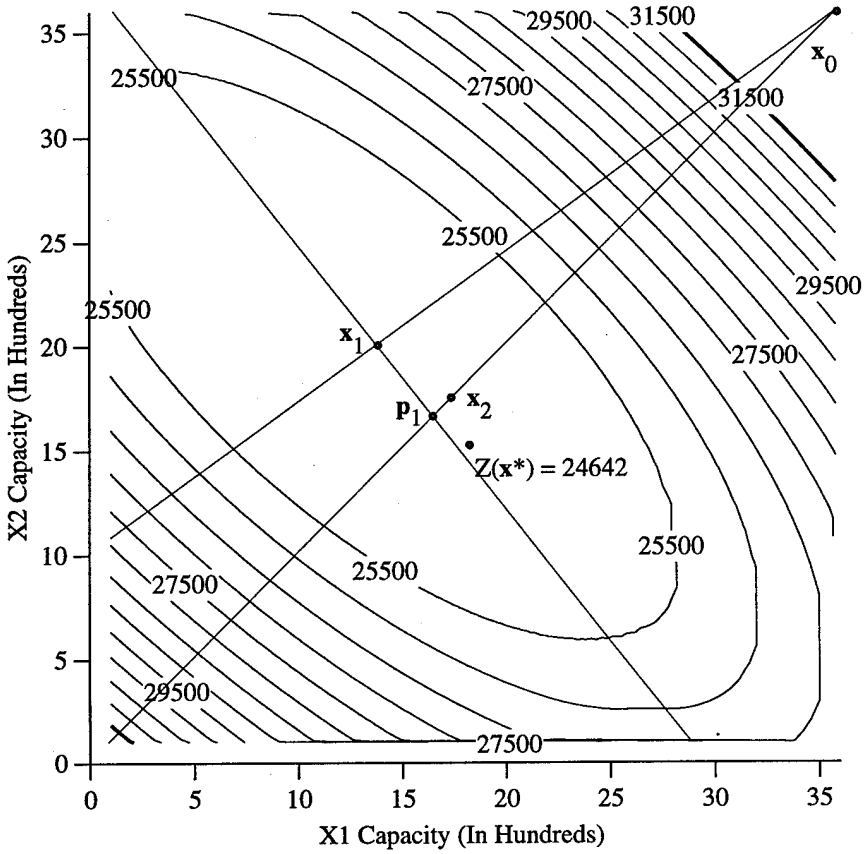


Figure 3.6. Example PARTAN Iteration for APL1P
(Lines Indicate Parallel Tangents)

TABLE 3.3
EXAMPLE PARTAN ITERATION FOR APL1P, SCENARIOS=1280, $Q = 6$

k	$\mathbf{x}$	Est. q^*	Act. q	R^2	$Z(\mathbf{x})^\dagger$
$\mathbf{x}_0$	3600, 3600	.613	.613	.990	32224.4
$\mathbf{x}_1$	1394, 1969	.513	.513	.996	24693.1
$\mathbf{p}_1$	1615, 1721	.539	.539	.990	24654.5
$\mathbf{x}_2$	1660, 1764	—	—	—	24655.1

† - $Z(\mathbf{x}^*) = 24642.3$, $\mathbf{x}^* = (1800, 1570)$

suggest a major source of error stems from the inaccuracy of the quadratic approximation (a phenomenon also present in the PROJECTED GRADIENT ALGORITHM).

3.3 LINEAR PROGRAMMING ALGORITHMS

3.3.1 Introduction

The cost of calculating $Z(\mathbf{x})$ described in the non-linear programming literature from which these methods are adapted usually presumes $Z(\mathbf{x})$ in a deterministic context. Unfortunately, in the present problem $h(\mathbf{x},\omega,\mathbf{T})$ compounds the computational expense by requiring statistical estimation of $Z(\mathbf{x})$. Therefore, search efficiency is no longer solely a matter of convergence, but one of response estimation as well. Consequently, this research looks into two broad areas of efficient estimation of $Z(\mathbf{x})$: *Linear Programming Algorithms* and *Variance Reduction*. Chapter 4 addresses variance reduction techniques, while the present section focuses on LP algorithmic improvements.

Obviously, the most straightforward method for solving z_{ik} would simply perform a standard primal or dual simplex algorithm for each new right-side vector $\omega_i - \mathbf{T}_i\mathbf{x}_k$, and perhaps retain the previous optimal basis in the hope of only

having to do a couple of pivots (if any). This optimization method, called the *OSL* option after the set of IBM library routines used to implement it, constitutes the first choice investigated by this research. It also forms the basis for comparing other alternatives for finding z_{ik} that begin with an assumption hinted at by the idea of minimal pivot requirements. These alternatives essentially employ the repetitive use of either an *Optimal Basis Set* (OBS) or an equivalent collection of *Optimal Dual Vectors* (ODV).

Specifically, the three proposed alternatives — the *OBS-COMPLETE, OBS-RESET, and ODV ALGORITHMS* — make the following crucial and fundamental assumption: *If a relatively small set of optimal bases characterize a subset of* $\boldsymbol{X}$*, then an algorithm faster than OSL can be developed by taking advantage of that set of optimal primal bases or its equivalent dual vectors.* Of course, the efficiency of the OBS and ODV methods depends on the size of the recourse problem (3.1b), the definition of the subset of $\boldsymbol{X}$, and the size of $\boldsymbol{X}$ itself. Yet even for large problems, there are cases where the OBS algorithm can prove advantageous. The following sections develop these concepts beginning with the following definitions.

Definition: Let $\mathbf{t}_{ik}$ be the i^{th} realization of $\omega_i - \mathbf{T}_i\mathbf{x}_k$; i.e., $\mathbf{t}_{ik} = \omega_i - \mathbf{T}_i\mathbf{x}_k$. Furthermore, let $\mathbf{t}_{ev} = E[\omega] - E[\mathbf{T}]\mathbf{x}$.

Definition. Let $\mathbf{P}_{ik}$ be the optimal recourse basis in (3.1b) for $\mathbf{t}_{ik} = \omega_i - \mathbf{T}i\mathbf{x}_k$; i.e., $(\mathbf{P}_{ik})^{-1}\mathbf{t}_{ik} = \mathbf{y}^*_{ik}$. Furthermore, let $P_{ev,k}$ be the optimal basis for $\mathbf{t}_{ev}$ and $\mathbf{x}_k$.

Definition. Let P represent a subset of the set of distinct positive cones containing every possible realization of $\omega - \mathbf{Tx}$.

Definition. Define L as the number of optimal bases in set P.

Definition. Let the random sample $\mathbf{t}_j \in \{\omega_j - \mathbf{T}_j\mathbf{x} : \mathbf{x} \in X\}$, for $j = 1, \ldots, J$.

Definition. Let the set T consist of sample realizations $\mathbf{t}_j$ which are *not* members of any positive cone defined by the elements of set P.

Definition. For an optimal set of bases P of size L to the recourse problem (3.1b), where optimal basis $\mathbf{P}_l \in P$, and $l = 1, \ldots, L$, let π_l be the optimal dual vector associated with $\mathbf{P}_l$, and let the set Π of size L consist of the corresponding set of optimal dual vectors π_l, $l = 1, \ldots, L$.

3.3.2 OBS-COMPLETE ALGORITHM

STEP 0. *(Initialization)* Select $\mathbf{x}$ and solve (3.1b) for $\mathbf{t}_{ev}$. Add optimal basis $\mathbf{P}_{ev}$ to P, associated optimal dual vector π_{ev} to Π, set $L = 1$ and remove elements in T. Notationally, $\mathbf{P}_{ev} = \mathbf{P}_1$.

STEP 1. Select sample size J and obtain random samples $\mathbf{t}_j$, $j = 1,\ldots, J$, using for each $\mathbf{t}_j$ a randomly selected $\mathbf{x} \in X$. Set non-optimal counter $n = 0$. Proceed with one of two options:

Frequency of Optimality: Set frequency of basis optimality counter $F_l = 0$, $l = 1, \ldots, L$. For *each* $\mathbf{t}_j$, $j = 1,\ldots, J$, and *each* optimal basis $\mathbf{P}_l \in \Pi$, $l =$

1,..., L, find $\mathbf{y}_{jl} = (\mathbf{P}_l)^{-1}\mathbf{t}_j$. If $\mathbf{y}_{jl} \geq \mathbf{0}$ set $F_l = F_l + 1$. If $\mathbf{y}_{jl} \ngeq \mathbf{0}$ $\forall$ $l = 1,..., L$, then put $\mathbf{t}_j$ in T and set $n = n + 1$. Index bases in P on l such that $F_1 \geq \ldots \geq F_l \geq \ldots \geq F_L$. Go to STEP 2.

Basic Coverage. For each $\mathbf{t}_j$, $j = 1,..., J$, find $\mathbf{y}_{jl} = (\mathbf{P}_l)^{-1}\mathbf{t}_j$. Terminate operation for current $\mathbf{t}_j$ upon finding first $\mathbf{y}_{jl} \geq \mathbf{0}$ and proceed with $\mathbf{t}_{j+1}$. If $\mathbf{y}_{jl} \ngeq \mathbf{0}$ $\forall$ $l = 1,..., L$, then put $\mathbf{t}_j$ in T and set $n = n + 1$. Go to STEP 2.

STEP 2. If $n = 0$ STOP. Otherwise, set $L = L + 1$. Select $\mathbf{t}_n$ from T. Solve (3.1b), add optimal basis $\mathbf{P}_n$ to P and optimal dual vector π_n to Π, remove elements in T, and repeat STEP 1.

3.3.3 OBS-RESET ALGORITHM

STEP 0. *(Initialization)* Clear optimal basis set P associated with previous first-stage vector $\mathbf{x}_{k-1}$. Solve for the first sample $\mathbf{t}_{1k}$ using OSL, and place the associated optimal basis $\mathbf{P}_1$ in P.

STEP 1. Select sample size J. For each following sample $\mathbf{t}_{jk}$, $j = 2, \ldots, J$, *first* check for a feasible solution from set P; i.e., where $(\mathbf{P}_l)^{-1}\mathbf{t}_{jk} \geq \mathbf{0}$, $l = 1, \ldots, L$ and $L \leq j$. If a feasible solution is found, then by Proposition 3.1 it is optimal. Otherwise, solve for $\mathbf{t}_{jk}$ using OSL, set $L = L + 1$, set $\mathbf{P}_L = \mathbf{P}_{jk}$, and place $\mathbf{P}_L$ in P. Repeat until completing sampling requirements.

3.3.4 ODV ALGORITHM

STEP 0. Select sample size J. For each following sample $\mathbf{t}_j$, $j = 1, \ldots, J$, by Proposition 3.2 find MAX$\{\pi_l(\omega_j - \mathbf{T}_j\mathbf{x}); \pi_l \in \Pi, l = 1, \ldots, L\}$.

3.3.5 Theoretical Development and Implementation

3.3.5.1 OBS-Complete and OBS-Reset

The idea of exploiting the existence of a small number of optimal bases for (3.1b) is not new — it motivates the sifting algorithm of Garstka and Rutenberg (1973); bunching (Wets 1983, 1988) and its extension by Haugland and Wallace (1988); the polyhedral cone decomposition algorithm for transportation problems proposed by Wallace (1986); and, is informally accepted as reasonable by current researchers in the field (Morton 1994b). However, unlike the sifting and bunching approaches which presume a lattice structure of discrete values for ω and $\mathbf{T}$, this dissertation proposes explicitly decomposing the region of interest into optimal bases (also referred to in the literature as positive cones, polyhedral cones, and decision regions) under more general conditions of continuous, non-independent random variables in ω and $\mathbf{T}$, without requiring any special structural properties other than what is already defined for (3.1). The following proposition from linear programming (see Murty 1983 or Bazaraa, Jarvis, and Sherali 1990) as applied in the context of the two-stage recourse problem (3.1) provides the theoretical underpinning for this approach by showing that for any variation of the right-side vector that remains within a positive cone defined by an optimal basis, then that basis is optimal for the perturbed value.

Definition: Let $\mathbf{t}_{rk}$ represent the r^{th} realization of $\omega_r - \mathbf{T}_r\mathbf{x}_k$; similarly let $\mathbf{t}_{sk}$ represent the s^{th} realization of $\omega_s - \mathbf{T}_s\mathbf{x}_k$.

Proposition 3.1: Unique Optimal Basis for Positive Cone

If $\mathbf{P}_{rk}$ is the optimal basis for $\mathbf{t}_{rk}$ in (3.1b), then for any $\mathbf{t}_{sk} \neq \mathbf{t}_{rk}$ and $\mathbf{t}_{sk} \in$ Pos($\mathbf{P}_{rk}$), $\mathbf{P}_{rk}$ is the optimal basis for $\mathbf{t}_{sk}$.

Proof: Let $\mathbf{t}_{sk} \in \text{Pos}(\mathbf{P}_{rk})$. By definition of positive cone there exists a vector $\mathbf{y}_{sk}$ such that $\mathbf{P}_{rk}\mathbf{y}_{sk} = \mathbf{t}_{sk}$, $\mathbf{y}_{sk} \geq \mathbf{0}$ whose primal solution value is $\mathbf{d}_P(\mathbf{P}_r,)^{-1}\mathbf{t}_{sk}$.

Since $\pi^*_{rk} = \mathbf{d}_P(\mathbf{P}_{rk})^{-1}$ dual feasibility remains unaffected, this gives the dual solution $\mathbf{d}_P(\mathbf{P}_{rk})^{-1}\mathbf{t}_{sk}$.

Since primal feasible and dual feasible solutions are equal, by fundamental duality $\mathbf{P}_{rk}$ is an optimal basis for $\mathbf{t}_{sk}$. ■

Obviously, the most efficient and useful result of Proposition 3.1 would be a small manageable set of optimal bases $P = \{\mathbf{P}_l, l = 1, 2, \ldots, L\}$, such that for any $\mathbf{x}_k \in \boldsymbol{X}$ and all realizations of ω and $\mathbf{T}$, $\mathbf{P}_k \in P$. In this case P contains every optimal basis (hence solution) for the entire space defined by $\mathbf{Ax=b}$, $\mathbf{x} \geq \mathbf{0}$, and for every realization of ω and $\mathbf{T}$; and, provides a computational improvement over an OSL-based algorithm in the following manner. Instead of performing the pivots of a revised primal or dual simplex for each new $\mathbf{t}_{sk}$, the OBS algorithm checks for a feasible solution by multiplying $\mathbf{t}_{sk}$ by the inverses of the bases in P. If a feasible solution is found (i.e., $(\mathbf{P}_l)^{-1}\mathbf{t}_{sk} = \mathbf{x}_{lsk} \geq \mathbf{0}$, $\mathbf{P}_l \in P$), then $z_{sk} = \mathbf{d}_P(\mathbf{P}_l)^{-1}\mathbf{t}_{sk}$. Furthermore, not *every* basis in P must be checked. Since Proposition 3.1 proves that *any* feasible basis in P is optimal, once a feasible basis is found the search may stop. (This would also suggest a rank ordering of the set P by *frequency of optimality*.) Finally, not every individual element of $\mathbf{x}_{lsk}$ must be checked; if x^i_{lsk}

(the i^{th} element of $\mathbf{x}_{lsk}$) is found to be negative, then $\mathbf{x}_{lsk}$ is automatically infeasible and thus the algorithm can immediately skip to the next basis $\mathbf{P}_{l+1}$.

This dissertation implements this approach using a separate program called the OBS-COMPLETE program. It precedes the second Monte Carlo simulation responsible for the experimental design (referred to as the *response surface approximation* (RSA) program) by preprocessing the recourse problem (3.1b) to find the complete optimal bases set P. The problem of finding all of the optimal bases for (3.1b) is equivalent to discovering, out of all the possible bases available from $\mathbf{W}$, the associated positive cones contained within the requirement space defined by $\omega - \mathbf{Tx}$, $\mathbf{x} \in X$. The OBS COMPLETE program accomplishes this task through an iterative process consisting of two key steps: (1) updating the set of optimal bases P and (2) uniformly sampling the recourse requirement space against the set of optimal bases P to determine the extent that *the sampled realizations* of $\omega - \mathbf{Tx}$ fall within the positive cones defined by current set P. Thus, the OBS-COMPLETE ALGORITHM essentially adds to and tests the set P until it possesses all optimal bases for the sampled realizations of $\omega - \mathbf{Tx}$. The RSA algorithm then uses the optimal bases set supplied by the OBS-COMPLETE program to estimate $Z(\mathbf{x})$ and ultimately derive the response surfaces of interest. (The ODV ALGORITHM also depends on the OBS-COMPLETE program to supply it with the dual vectors associated with the primal optimal bases.)

Although *in theory* such an optimal basis set can be found for any problem, as a *practical* matter only relatively small recourse problems allow such complete 'coverage' of the feasible region. Furthermore, at some point the size of the set P, in combination with the frequency distribution of optimality among its

elements, will offset the computational advantages just described. Yet even for larger problems the OBS approach can still prove helpful if the region of interest is a small subset of $\boldsymbol{X}$ — *even to the point of being distinct vector* $\boldsymbol{x}$. Therefore, this research applies this basic technique in two ways:

1. *OBS-COMPLETE.* This method defines the region of interest as $\boldsymbol{X}$; in other words, complete coverage of the recourse requirements space ω - $\mathbf{Tx}$, $\mathbf{x} \in \boldsymbol{X}$, can be provided by a small set of optimal bases P.

2. *OBS-RESET.* This method defines the subset of $\boldsymbol{X}$ to be a single vector $\mathbf{x}_k$. In effect, this requires the optimal basis set P_k to be cleared (or 'reset') for each new vector $\mathbf{x}_{k+1}$. Although clearly not as efficient as the OBS-COMPLETE method, for larger recourse problems with numerous stochastic variables in ω and a bigger dimensioned $\mathbf{x}$, such an approach can still offer computational advantages over an OSL-based estimation method.

Unlike the OBS-COMPLETE or ODV ALGORITHMS, OSL-RESET does not require a preprocessed basis set. Therefore, the response surface analysis algorithm RSA directly incorporates the OBS-RESET ALGORITHM along with the OSL option. Finally, because of the additional computational and storage expense, OBS-RESET does not store an associated set of dual vectors since they do not provide any additional information on the estimate of $Z(\mathbf{x})$.

3.3.5.2 ODV

The third basic approach to improving the efficiency of estimating $Z(\mathbf{x})$ from a linear programming algorithmic perspective uses the set of dual vectors associated with the optimal bases set P. Using an idea suggested by Morton (1995) and duality results from linear programming theory (see Murty 1983 or Bazarra, Jarvis, and Sherali 1990), the following proposition establishes the ODV technique.

Proposition 3.2: Dual Optimality of the Recourse Problem

Given a finite set of optimal bases P which contains at least one optimal basis for any realization of $\omega - \mathbf{Tx}$ *to the primal recourse problem* MIN $\mathbf{dy}$ *s.t.* $\mathbf{Wy} = \omega - \mathbf{Tx}$, $\mathbf{y} \geq \mathbf{0}$, *with a corresponding set of optimal dual vectors* Π *for the dual recourse problem* MAX $\pi(\omega - \mathbf{Tx})$, *s.t.* $\pi\mathbf{W} \leq \mathbf{d}$, π *unrestricted; then, for any realization of* $\omega - \mathbf{Tx}$ *the optimal solution* $\mathbf{dy}^* = \text{MAX}\{\pi_l(\omega - \mathbf{Tx}), l = 1, 2, \ldots, L; \pi_l \in \Pi\}$.

Proof: By assumption any realization of $\omega - \mathbf{Tx}$ has the property of possessing an optimal basis $\mathbf{P}^* \in P$ and its corresponding optimal dual vector $\pi^* \in \Pi$.

Since the dual constraint $\pi\mathbf{W} \leq \mathbf{d}$ remains constant, every $\pi_l \in \Pi$ is feasible for any realization of $\omega - \mathbf{Tx}$.

By strong duality $\mathbf{dy}^* = \pi^*(\omega - \mathbf{Tx})$ at optimality.

By weak duality $\mathbf{dy}^* \geq \pi_l(\omega - \mathbf{Tx})$, $\pi\mathbf{W} \leq \mathbf{d}$. Since by Step 1 $\pi^* \in \Pi$ and by Step 2 every $\pi_l \in \Pi$ is feasible, the optimal solution must be

MAX$\{\pi_l(\omega - \mathbf{Tx}), l = 1, 2, \ldots, L; \pi_l \in \Pi\}$; otherwise, it would contradict Step 3. ■

The ODV ALGORITHM directly implements Proposition 3.2 by simply multiplying the current recourse right-side vector $(\omega_j - \mathbf{T}\mathbf{x}_k)$ against all dual vectors $\pi_l \in \Pi$, and selecting the highest resulting product as the optimal solution. Unlike the OBS method, where the search through the optimal bases set P stops after finding a feasible solution, ODV must check every dual vector π in Π in order to guarantee an optimal solution. For this reason, a dual vector counterpart to the OBS-RESET option is not possible since primal feasibility cannot be determined directly from Π. In other words, an ODV ALGORITHM can only be used where a complete dual vector set Π (and associated optimal bases set P) for the subset of X has been assembled.

However, the ODV ALGORITHM may still offer computational advantages over either the OSL or OBS-COMPLETE versions for cases where a complete dual vector set Π can be constructed. For a recourse problem with m constraints, each basis requires no more than m^2 multiplications and $m \cdot (m - 1)$ additions; thus, where P contains L bases would (worst case) require $L \cdot (2m^2 - m)$ arithmetic operations. By contrast, each dual vector requires m multiplications and $(m - 1)$ additions, and for Π containing L dual vectors the ODV ALGORITHM requires $L \cdot (2m - 1)$ operations. Clearly, most cases the OBS algorithm will not realize the upper bound on the number of operations due to compact storage for sparse matrices; the potential to skip to the next basis upon finding a negative element in the current solution vector; and, the ability to stop after finding a feasible basis

and not search the remaining elements in set P. However, comparing the ratio of the worst case bound for OBS-COMPLETE with the known arithmetic operations requirement of ODV shows that the combined effects of the above-mentioned items would have to reduce the computational demands of the OBS-COMPLETE ALGORITHM by a factor of *m on average* in order to for it to compete with the ODV ALGORITHM. Therefore, the ODV ALGORITHM will most likely outperform the OBS-COMPLETE ALGORITHM except for small sets P and Π with a very skewed distribution of the frequency of optimality among the bases in P.

3.3.6 OBS-Based Results for APL1P

The APL1P problem's low dimensionality and small probability space make it a good candidate for the OBS-COMPLETE ALGORITHM. The major difficulty lies in defining the requirements space for ω -**Tx**; again, the range of **Tx** posing the biggest hurdle. By inspection it is clear that **x** is unbounded from above and each element must be greater than 1. However, it is equally clear that a *practical* bound exists based on the upper limits of demand ω; indeed, the highest possible demand from either supply 1 or 2 would be where all elements ω^j are 1200 and only one supply node (1) satisfies the total demand of 3,600 (an event whose probability is 6.75×10^{-5}). Therefore, for this problem the range of x^i is [1,3600].

Another complicating factor concerns the stochastic elements in **T**. In the case of APL1P they have no effect on the *range* of the requirement space since the highest possible multiple in **T** renders any $x^i > 3600$ meaningless. (By counter example, if the highest multiple in **T** was .5, then $x^i \in [1,7200]$.) This means that

4 independent discrete states for each ω^j and ξ^{11}, and 5 for ξ^{22} results in 1280 independent possible realizations of the stochastic parameters, or *scenarios*. Obviously a random sampling of these scenarios would provide a faster evaluation of $Z(\mathbf{x})$, and the topic is taken up in Chapter 4. However, for the search techniques and optimization options discussed in the present chapter all 1280 scenarios are evaluated (i.e., every $Z(\mathbf{x})$ represents the true expected value of the objective function for (3.1a)).

Using the data provided in Figure 3.1, the OBS-COMPLETE ALGORITHM found 13 bases (and associated dual vectors) that thoroughly describe the requirements space for APL1P. Table 3.4 provides the search data and computational results, and Table 3.5 follows with estimates on the frequency of optimality for each basis/dual vector. For these and subsequent results, the reported computational times represent the operating system's estimate for the CPU time required to execute the algorithm — it *does not* include system I/O time or overhead (IBM 1992). All examples and problems were written and compiled in FORTRAN 90, and run on an IBM RS/6000 Model 320 under AIX 3.2.

Tables 3.6 through 3.8 compare computational times of the three optimization options for each of the three search techniques discussed in Sections 3.2.2 through 3.2.7. Each Table contains two independent search paths, with the first entry corresponding to the example search path shown in Figures 3.3, 3.4, and 3.6. In all cases the OBS-COMPLETE and ODV options are significantly faster than the OSL — enough so that at least an order of magnitude difference in speed occurred, thus justifying the upfront cost in finding the optimal basis/dual vector set.

TABLE 3.4
OBS-COMPLETE RESULTS FOR APL1P

Sample Size (ω - **Tx**)	Random # Seed	All Bases / First Optimal*	# Opt. Bases/ Dual Vectors	CPU Time (secs)
1000	873946	All	10	3.81
5000	4209175	All	11	5.80
25000	3366149	All	12	29.34
200000	66231850	First	13	139.55

* - 'First Optimal' Option Skips Any Remaining Bases After Finding First Feasible, Whereas 'All Bases' Checks Every Basis in P for each Sample (ω - **Tx**)

TABLE 3.5
FREQUENCY OF BASIS OPTIMALITY
(BASED ON 4th RUN FROM TABLE 3.4)

Basis ID #	1	2	3	4	5	6	7	8	9	10	11	12	13
Freq. of Optimality	.19	.18	.12	.11	.10	.10	.08	.06	.04	.04	.00	.00	.00
Cumulative Freq.*	.19	.36	.48	.58	.68	.78	.86	.92	.96	1.0	1.0	1.0	1.0
Lowest **x** Sampled	(0.85, 0.00)												
Highest **x** Sampled	(3596.36, 3561.52)												

* - May Not Add Due to Roundoff Error

TABLE 3.6
COMPUTATION TIMES OF OSL/OBS/ODV OPTIONS OF GEOMETRIC SIMPLEX ALGORITHM FOR TWENTY SIMPLEX ITERATIONS OF APL1P (IN SECONDS)

Starting Simplex Parameters	OSL	OBS	ODV
$\mathbf{x}_1$ = (1529,1625), $\mathbf{x}_2$ = (2437,2681), $\mathbf{x}_3$ = (2502,3462) Random Seed = 440908571, 20 Total Vertices	274.07	19.77	16.62
$\mathbf{x}_1$ = (1529,1625), $\mathbf{x}_2$ = (725,718), $\mathbf{x}_3$ = (3203,3561) Random Seed = 707446, 20 Total Vertices	171.64	13.53	10.39

TABLE 3.7
COMPUTATION TIMES OF OSL/OBS/ODV OPTIONS OF PROJECTED GRADIENT ALGORITHM FOR APL1P (IN SECONDS)

Search Parameters	OSL	OBS	ODV
$\mathbf{x}_1$ = (3600, 3600), Q = 6, Iterations = 10	612.22	25.09	23.05
$\mathbf{x}_1$ = (2, 2), Q = 8, Iterations = 10	807.39	32.31	30.29

TABLE 3.8
COMPUTATION TIMES OF OSL/OBS/ODV OPTIONS OF PARTAN ALGORITHM FOR APL1P (IN SECONDS)

Search Parameters	OSL	OBS	ODV
$\mathbf{x}_1 = (3600, 3600)$, $Q = 6$, Iterations = 4	340.62	17.48	13.09
$\mathbf{x}_1 = (2, 2)$, $Q = 6$, Iterations = 4	346.78	17.71	16.28

Chapter 4

Methodology: Statistical Analysis

4.1 INTRODUCTION

Restating (3.1), this dissertation restricts its focus to the class of two-stage stochastic linear programming problems with recourse of the form

$$\text{MIN } Z(\mathbf{x}) = \mathbf{cx} + E[h(\mathbf{x},\omega,\mathbf{T})], \text{ s.t. } \mathbf{Ax} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0} \tag{3.1a}$$

$$h(\mathbf{x},\omega,\mathbf{T}) = \text{MIN } \mathbf{dy}, \text{ s.t. } \mathbf{Wy} = \omega - \mathbf{Tx}, \ \mathbf{y} \geq \mathbf{0} \tag{3.1b}$$

where a finite mean and variance exist for each component of ω and $\mathbf{T}$. The previous chapter's focus on search techniques and optimization methods falls within the traditional framework of solving stochastic recourse problems of the form (3.1) by synthesizing previously disparate ideas into a unified and unique methodology for solving (3.1a). However, these computational advances merely improve upon an existing analytical paradigm without offering additional insight into the problem. When viewed from a *statistical perspective*, though, (3.1a) presents a completely different challenge to the analyst and decision-maker. This challenge is not simply the complications in finding the optimal solution $\mathbf{x}^*$ posed by the random variables present in ω and $\mathbf{T}$ (although this is certainly an issue and a subject of much research). Instead, the nature of the problem itself is altered in the following fundamental way: *Any realistic and useful answer to (3.1a) becomes less one of strict optimality and more an issue of (1) sensitivity and (2)*

variability. This chapter explores this new analytical approach starting with a review of the following primary definitions from Chapter 3.

Definition. Let $X = \{\mathbf{x} : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$.

Definition Let $\mathbf{x}^*$ represent the optimal solution that minimizes $Z(\mathbf{x})$; i.e., $Z(\mathbf{x}^*) = \text{MIN } Z(\mathbf{x})$. Define the *region of optimality* as the set $\{\mathbf{x}' : Z(\mathbf{x}') \leq Z(\mathbf{x}^*) + \varepsilon, \varepsilon > 0, \mathbf{x}' \in X\}$; i.e. those feasible solutions $\mathbf{x}'$ whose objective values $Z(\mathbf{x}')$ are near-optimal (as defined by ε).

Definition. Let $z_{ik} = \mathbf{c}\mathbf{x}_k + h(\mathbf{x}_k, \omega_i, \mathbf{T}_i)$; i.e., z_{ik} represents the objective value given $\mathbf{x}_k$ for the i^{th} realization of ω and $\mathbf{T}$. Further define the independent random variable z_k distributed as $\mathbf{c}\mathbf{x}_k + h(\mathbf{x}_k, \omega, \mathbf{T})$, where $E[z_k] = Z(\mathbf{x}_k)$.

The first type of inquiry involves an area of analysis called *response surface methodology* that deals with the *shape* of the response function $Z(\mathbf{x})$, and is motivated by the following observations. First, the traditional optimal answer minimizes the first moment of the recourse function, even though $\mathbf{x}^*$ may have undesirable characteristics that near optimal solutions $\mathbf{x}'$ may not share. If the region of optimality is 'flat', however, then the difference between $Z(\mathbf{x}^*)$ and $Z(\mathbf{x}')$ may be small enough to justify eliminating the unwanted attributes of $\mathbf{x}^*$. Using APL1P as an example, there does indeed exist an optimal answer ($\mathbf{x}^* = (1801.9, 1571.4)$, $Z(\mathbf{x}^*) = 24642.3$); yet, a quick glance at the figures and data in Chapter 3 reveals a region of extremely low sensitivity of $Z(\mathbf{x})$ to changes in $\mathbf{x}$.

For instance, in Table 3.2 where $\mathbf{x}_2$ = (1394,1969), $Z(\mathbf{x}_2)$ = 24693.1, a mere .2% increase in the expected optimal objective function value occurs for a -22.6% and 25.3% change in x^1 and x^2, respectively. Similarly, there exists a (infinitely) large number of near-optimal solutions $\mathbf{x}'$ that roughly follow a line of -1 slope running through the center of the ellipsoid. Therefore, unless an analyst could claim that (3.1) truly captures *all* relevant objectives and constraints, or that no subjective criteria influences the decision-maker, then some measure of solution sensitivity becomes necessary.

Extending this idea, the second — and equally important — aspect of sensitivity analysis concerns knowing where *not* to move. For instance, proceeding an equal geometric distance in a direction orthogonal to the vector from $\mathbf{x}^*$ to $\mathbf{x}_2$ in APL1P gives $\mathbf{x}_o$ = (2199.5, 1979.3) and $Z(\mathbf{x}_o)$ = 25244.3. This produces a 2.44% increase in objective function value (over 11 times as sensitive as compared to $\mathbf{x}_2$) for a 22.1% and 26.0% change in x^1 and x^2, respectively. Generalizing such an approach, this chapter will show how special canonical transformations of the original response surface furnish such minimal and maximal ridge analysis for n-dimensional problems, thus providing a very important and basic tool for characterizing $Z(\mathbf{x})$.

Finally, several aspects of the distribution of $h(\mathbf{x},\omega,\mathbf{T})$ present another major reason why settling for the solution $\mathbf{x}^*$ can be deceptive. First, although $Z(\mathbf{x}^*)$ by definition provides the minimum expected value, decision-makers are often risk-averse — in short, instead of minimizing *expected value* they may wish to avoid the *worst-case* scenario. (For example, an optimal $\mathbf{x}^*$ may be less desirable than a near-optimal solution $\mathbf{x}'$ if $\text{VAR}[h(\mathbf{x}',\omega,\mathbf{T})] << \text{VAR}[h(\mathbf{x}^*,\omega,\mathbf{T})]$.)

Second, as the literature clearly suggests, the distribution of $h(\mathbf{x},\omega,\mathbf{T})$ itself is a function of $\mathbf{x}$, and may not necessarily follow a symmetric distribution (preliminary research found several empirical examples of highly skewed, asymmetric distributions). However, any unstated assumptions on the part of the decision-maker regarding the distributional form of $h(\mathbf{x},\omega,\mathbf{T})$ can be misleading when used in conjunction with *just* $E[h(\mathbf{x},\omega,\mathbf{T})]$. Therefore, non-parametric analyses, such as tolerance limits and quantile-based statistics (such as the median), can provide very practical information to supplement expected values when comparing $\mathbf{x}'$ to $\mathbf{x}^*$, or for simply understanding the underlying behavior of $h(\mathbf{x}^*,\omega,\mathbf{T})$.

The third aspect of $h(\mathbf{x},\omega,\mathbf{T})$ — variance — also influences both the search techniques discussed in the previous chapter and the validity of the response surface approximation of $Z(\mathbf{x})$. This occurs because the tremendously large number of scenarios associated with many recourse problems requires a sampling of the probability space of ω - $\mathbf{Tx}$, whose corresponding estimates of the expected values reduces the accuracy of the experimental design and response surface analysis. Under these conditions, *variance reduction* becomes especially important; it not only increases computational efficiency for the search techniques, but also reduces the adverse impact of sample variance — which in turn improves the accuracy of the polynomial approximation of the response $Z(\mathbf{x})$. Although some work has been done on variance reduction in the context of improved estimators, the literature does not offer prior research on these two important aspects of the problem from the perspective of the decision-maker.

This dissertation proceeds from the premise that these two characteristics — sensitivity and variance — provide the most important insights into the behavior of (3.1a). This chapter presents the techniques used to accomplish these two goals under the following topics: *Variance Reduction, Experimental Design, Response Surface Analysis,* and *Distribution Analysis.*

4.2 VARIANCE REDUCTION

4.2.1 Introduction

When the probability models representing the stochastic elements of ω and $\mathbf{T}$ contain a large number of discrete realizations, or take a continuous form, then an unbiased estimate $\hat{Z}_s(\mathbf{x})$ of the true population mean $Z(\mathbf{x})$ through sampling of the population becomes necessary. Recalling the definition of z_{ik}, the first unbiased estimator of $Z(\mathbf{x}_k)$ this dissertation uses employs random sampling of size I where

$$\hat{Z}_{RS}(\mathbf{x}_k) = (I)^{-1} \cdot \sum_{i=1}^{I} z_{ik}. \tag{4.1}$$

It follows directly that since $\hat{Z}_{RS}(\mathbf{x}_k)$ itself is a random variable, it has a variance $\text{VAR}[\hat{Z}_{RS}(\mathbf{x}_k)]$ defined as

$$\text{VAR}[\hat{Z}_{RS}(\mathbf{x}_k)] = \text{VAR}[(I)^{-1} \cdot \sum_{i=1}^{I} z_{ik}] = (I)^{-2} \cdot [\sum_{i=1}^{I} \text{VAR}(z_{ik})] = \frac{\sigma^2_k}{I} \tag{4.2}$$

where an unbiased estimate $S^2_k(I)$ of σ^2_k can be found using the relation

$$S^2_k(I) = (I\text{-}1)^{-1} \cdot \sum_{i=1}^{I} [z_{ik} - \hat{Z}_{RS}(\mathbf{x}_k)]^2 \tag{4.3}$$

(Law and Kelton 1991). However, the problem with using the random sample estimate $\hat{Z}_{RS}(\mathbf{x}_k)$ for $Z(\mathbf{x}_k)$ lies precisely with (4.2) — such variability implies error. The ramifications of estimator variance on the search process have already been addressed (e.g., Barton and Ivey (1991) and the false convergence of the GEOMETRIC SIMPLEX ALGORITHM); however, estimator variability can also profoundly affect the accuracy and validity of the response surface estimates. For example, if a single point estimator $\hat{Z}_{RS}(\mathbf{x}_k)$ lying in the 'flat' region of near-optimality as part of an experimental design underestimates the true response $Z(\mathbf{x}_k)$, then the resulting polynomial approximation could easily assume the surface curves *downward* in the direction represented by the errant design point. This in turn would produce a response surface shape resembling a saddle where the true shape is known to be convex (see Box and Draper 1987).

As the name implies, variance reduction techniques (VRTs) attack this problem by trying to reduce the value of (4.2) *in an efficient manner*. Obviously, the sample variance can be decreased by increasing the sample size I, but this becomes computationally prohibitive for large recourse problems. The preferred approach reduces sample variance with the same I, or equivalently gives (4.2) the same magnitude with fewer simulations. There are a wide variety of VRTs; Law and Kelton (1991) provide an excellent description of the major VRT categories, and suggest secondary references Nelson (1987) and Wilson (1984) for more comprehensive reviews. However, as noted in Chapter 2 the application of VRTs to the recourse problem in the literature remains limited, requiring much additional work on both increased individual point estimator efficiency and its

larger effects on response surface approximation. This research contributes to this area by exploring two areas of VRTs: *Control Variates* and *Latin Hypercube.*

4.2.2 Control Variates

In very general terms, most VRTs attempt to reduce the sample variance by correlating some internal aspect of the simulation to the response being estimated. For instance, Common Random Numbers (CRNs) compare alternative configurations under the same random number stream; Antithetic Variables (AVs) use complementary random numbers under the assumption that the opposing pairs are negatively correlated; and, Conditioning Estimation (CE) employs known analytical values in lieu of estimates where possible (Law and Kelton 1991). For reducing the sample variance in the recourse context, Control Variates (CVs) is an attractive technique for the reasons outlined in Law and Kelton (1991):

1. *Correlation.* Contrary to CRN and AV methods, CVs work with either positive or negative correlation.

2. *Simplicity.* CVs do not require separate runs like CRNs, or synchronized replication like AVs.

3. *Effectiveness.* If *any* correlation exists between a control variate and z_{ik}, then CVs will reduce the sample variance.

4. *Efficiency.* Using internal CVs does not appreciably increase the computational requirements of the simulation.

Finally — and most importantly — this dissertation contends that the recourse problem presents an evident set of random variables that strongly support CVs as an effective and efficient VRT candidate.

Following Law and Kelton's (1991) presentation, CVs assume that a random variable X with a known expectation E[X] is correlated (positively or negatively) with the simulation response Y, where E[Y] is unknown and thus estimated. Intuitively, if during a simulated observation X is greater than E[X], then the resulting response Y should also be greater (positive correlation) or lesser (negative correlation) than E[Y] and adjusted accordingly. This relationship can be expressed mathematically as

$$Y_C = Y - b(X - E[X]) \tag{4.4}$$

where Y_C represents the controlled, unbiased estimator for E[Y]; and, b is a constant whose sign corresponds to the correlation between Y and X, and whose value quantifies the adjustment. Taking the variance of (4.4) gives

$$\text{VAR}[Y_C] = \text{VAR}[Y] + b^2\text{VAR}[X] - 2b\text{COV}[Y;X] \tag{4.5}$$

from which it immediately follows that $\text{VAR}[Y_C] < \text{VAR}[Y]$ if and only if

$$2b\text{COV}[Y;X] > b^2\text{VAR}[X]. \tag{4.6}$$

Regarding $\text{VAR}[Y_C]$ as a function of b and setting its derivative to zero gives

$$b^* = \frac{\text{COV}[Y;X]}{\text{VAR}[X]} \tag{4.7}$$

whose substitution for b in (4.5) yields

$$\text{VAR}[Y_{C^*}] = (1 - \rho^2_{YX}) \cdot \text{VAR}[Y] \tag{4.8}$$

or the minimum variance adjusted estimator Y_{CV^*}, where ρ_{YX} is the correlation between Y and X. From (4.8) it follows that if Y and X are at all correlated then $\text{VAR}[Y_{CV^*}] < \text{VAR}[Y]$; and, the higher the correlation the greater the variance reduction (Law and Kelton 1991; also see Kleijnen 1974, Lavenberg and Welch 1981, Nelson 1987, and Nelson 1990).

The relationship described in (4.4) is present in the recourse problem (3.1) by observing that the estimator z_{ik} should be correlated with one or more of the random components of ω and $\mathbf{T}$.

Definition. Define the P-dimensional vector $\mathbf{v}$ whose elements v^p, $p = 1, \ldots, P$, represent selected components of ω or $\mathbf{T}$ with known $E[v^p]$ and $\text{VAR}[v^p]$.

Definition. Let $\mu^p = E[v^p]$ and define the P-dimensional vector $\mu = [\mu^1, \mu^2, \ldots, \mu^P]$.

Definition. Let $b^*_{vp} = \dfrac{\text{COV}[z_k; v^p]}{\text{VAR}[v^p]}$.

The previous definitions allow for an unbiased minimum adjusted estimator for the i^{th} realization of v^p

$$z^*_{ik} = z_{ik} - b^*_{vp}(v^p_i - \mu^p), \tag{4.9}$$

where the unbiased CV estimator of $Z(\mathbf{x}_k)$ is

$$\hat{Z}_{CV}(\mathbf{x}_k) = (I)^{-1} \cdot \sum_{i=1}^{I} z_{ik}^{*} \tag{4.10}$$

and where $\text{VAR}[\hat{Z}_{CV}(\mathbf{x}_k)] \leq \text{VAR}[\hat{Z}_{RS}(\mathbf{x}_k)]$ (Law and Kelton 1991). While clearly a correlation almost certainly exists between z_{ik} and the random variables in ω and $\mathbf{T}$, *which* stochastic element — or combination of elements — in ω and $\mathbf{T}$ provides the greatest variance reduction as a control variate is not at all apparent. Answering this question raises the following issues of CV selection, bias, and non-stationarity.

4.2.2.1 Control Variate Selection

Following Lavenberg and Welch (1981), all previous scalar CV formulations can be extended to include multiple CVs; i.e., replacing v^p with the vector of controls $\mathbf{v}$ gives

$$z_{ik}^{*} = z_{ik} - \boldsymbol{b}_{\mathbf{v}}^{*}(\mathbf{v}_i - \mu) \tag{4.11}$$

where

$$\boldsymbol{b}_{\mathbf{v}}^{*} = \sigma_{\mathbf{vz}} \cdot [\Sigma_{\mathbf{v}}]^{-1}, \tag{4.12}$$

$\Sigma_{\mathbf{v}}$ is the covariance matrix for $\mathbf{v}$, $\sigma_{\mathbf{vz}}$ is the covariance vector for $\mathbf{v}$ and z_{ik}, the elements of the P-dimensional vector $\mathbf{v}_i$ are the $i^{\underline{th}}$ realization of the random variable v_i^p, $p = 1, \ldots, P$, and the elements of the vector $\boldsymbol{b}_{\mathbf{v}}^{*}$ characterize the optimum coefficients for maximum variance reduction. However, in so doing the loss factor

$$\frac{I-2}{I-P-2} \tag{4.13}$$

occurs due the estimation requirements of (4.12) discussed shortly (Lavenberg and Welch 1981; also see Law and Kelton 1991, and Rubenstein and Marcus 1985).

While (4.13) encourages a parsimonious selection of multivariate CVs, it still does not decide *which* random variables in ω or **Tx** to select. This research addresses these topics in the following manner. For v^p select a component of ω or **T** assumed to be most closely correlated to z_{ik} based on empirical observations from search results, preliminary screening designs, or subjective knowledge of the underlying system. Reduce the loss factor described in (4.13) and minimize the computational requirements implicit in estimating (4.12) by letting $P \leq 3$. This heuristic follows the findings of Rubenstein and Marcus (1985) suggesting that too large a P can over-correct the controlled estimator.

4.2.2.2 Bias Estimators

While $\text{VAR}[v^p]$ is known, $\text{COVAR}[z_k;v^p]$ is not (the same holds true for Σ_v and σ_{vz}, respectively, in the multiple CV case); therefore, the covariance relationship must be estimated. Unfortunately, such estimation creates a biased estimator of $Z(\mathbf{x})$ as Law and Kelton (1991) show in the scalar case. Substituting the estimator $\hat{b}^*_{vP}$ for b^*_{vP} in (4.9), where

$$\hat{b}^*_{vP} = \hat{\text{COVAR}}[z_k;v^p] \tag{4.14}$$

gives the estimator for $\text{COVAR}[z_k,v^p]$ as

$$\hat{\text{COVAR}}[z_k;\mathbf{v}^p] = (I-1)^{-1} \cdot \sum_{i=1}^{I}(z_{ik} - \hat{Z}(\mathbf{x}_k))(v_i^p - \hat{\mu}^p) \tag{4.15}$$

by letting $\hat{\mu}^p$ (the sample estimate for μ^p) be defined as

$$\hat{\mu}^p = (I)^{-1} \cdot \sum_{i=1}^{I} v_i^p. \tag{4.16}$$

In turn, this gives the new observation as

$$\hat{z}_{ik}^* = z_{ik} - \hat{b}_{\mathbf{v}P}^*(v_i^p - \mu^p). \tag{4.17}$$

Consequently $\hat{z}_{ik}^*$, $i = 1, \ldots, I$, are no longer independent since they all contain $\hat{b}_{\mathbf{v}P}^*$. Therefore, expectations cannot be taken across (4.17) in the same manner as for (4.9) (Law and Kelton 1991).

Like its scalar counterpart $\boldsymbol{b}_{\mathbf{v}}^*$ must also be estimated, and therefore its estimator will also be biased for the same reasons. Following Lavenberg, Moeller, and Welch (1982) a biased estimate associated with $\boldsymbol{b}_{\mathbf{v}}^*$ follows from using the multiple CV version of (4.14),

$$\hat{\boldsymbol{b}}_{\mathbf{v}}^* = \hat{\sigma}_{\mathbf{v}z} \cdot [\hat{\Sigma}_{\mathbf{v}}]^{-1}, \tag{4.18}$$

where $\hat{\Sigma}_{\mathbf{v}}$ is the sample covariance matrix whose $qp^{\underline{th}}$ component (rows indexed on $q = 1, \ldots, P$, and columns indexed on $p = 1, \ldots, P$) is

$$\hat{\Sigma}_{\mathbf{v}}^{qp} = (I-1)^{-1} \cdot \sum_{i=1}^{I}(v_i^q - \hat{v}^q)(v_i^p - \hat{\mu}^p) \tag{4.19}$$

and $\hat{v}^q$ represents the sample estimate for v^q

$$\hat{v}^q = (I)^{-1} \cdot \sum_{i=1}^{I} v_i^q. \tag{4.20}$$

Similarly, the sample covariance vector $\hat{\sigma}_{vz}$ uses the relation

$$\hat{\sigma}_{vz}^p = (I - 1)^{-1} \cdot \sum_{i=1}^{I} (z_{ik} - \hat{Z}_s(\mathbf{x}_k))(v_i^p - \hat{\mu}^p) \tag{4.21}$$

to estimate σ_{vz} (Lavenberg, Moeller, and Welch 1982). One alternative to eliminating biased estimators due to dependent estimates of b_{vp}^* (4.14) or $\boldsymbol{b}_v^*$ (4.18) with z_k would be to estimate $\hat{b}_{vP}^*$ or $\hat{\boldsymbol{b}}_v^*$ using separate data; however, that would increase the number of samples for $\mathbf{x}_k$. Instead, a more efficient approach uses *jackknifing* (or *generalized splitting*) (Kleijnen 1974; Lavenberg, Moeller, and Welch 1982; and, Miller 1974).

This research implements a simple jackknife estimate using the following procedure as described by Lavenberg, Moeller, and Welch (1982).

Definition. Let $\bar{Z}(\mathbf{x}_k) = (I)^{-1} \cdot \sum_{i=1}^{I} z_{ik}$.

Definition. Let $\bar{\mathbf{v}} = (I)^{-1} \cdot \sum_{i=1}^{I} \mathbf{v}_i$.

Definition. Let the set $J_k = \{z_{1k}, \ldots, z_{Ik}; v_{1k}^1, \ldots, v_{Ik}^1; \ldots; v_{1k}^P, \ldots, v_{Ik}^P\}$ represent all observations of z_{ik} and $\mathbf{v}_{ik}$, $i = 1, \ldots, I$ for $\mathbf{x}_k$.

Temporarily dropping the k subscript for convenience, the biased estimator for $Z(\mathbf{x})$ becomes

$$\tilde{Z}(\mathbf{x}) = \bar{Z}(\mathbf{x}) - \hat{\boldsymbol{b}}_v^*(\bar{\mathbf{v}} - \mu_v). \tag{4.22}$$

Partitioning the I observations of set J into G sets of equal length H such that $I = GH$, let index g denote the set of observations

$$\{\mathbf{z}_g,\mathbf{v}_g\} = \left\{ \begin{array}{cccc} z_{g1} & v^1_{g1} & \cdots & v^p_{g1} \\ z_{g2} & v^1_{g2} & \cdots & v^p_{g2} \\ \vdots & \vdots & \vdots & \vdots \\ z_{gH} & v^1_{gH} & \cdots & v^p_{gH} \end{array} \right\}, \; g = 1, \ldots, G \tag{4.23}$$

where, if $\{\mathbf{z}_i,\mathbf{v}_i\} \subseteq J$, $\{\mathbf{z}_l,\mathbf{v}_l\} \subseteq J$, and $i \neq l$, then $\{\mathbf{z}_i,\mathbf{v}_i\} \cap \{\mathbf{z}_l,\mathbf{v}_l\} = \varnothing$. Let $\tilde{Z}_g(\mathbf{x})$ represent an estimator of $Z(\mathbf{x})$ using the same formulation of (4.22) *without* $\{\mathbf{z}_g,\mathbf{v}_g\}$; i.e., $\tilde{Z}_g(\mathbf{x})$ estimates $Z(\mathbf{x})$ using all observations from J *except* the subset $\{\mathbf{z}_g,\mathbf{v}_g\}$. Similarly, let $\hat{\boldsymbol{b}}^*_{\mathbf{v}_g}$ represent the estimator $\hat{\boldsymbol{b}}^*_{\mathbf{v}}$ calculated from all the data in J *except* $\{\mathbf{z}_g,\mathbf{v}_g\}$ using (4.18). Then G distinct estimates of $Z(\mathbf{x})$ can be found using the relationship

$$\tilde{\phi}_g = G\tilde{Z}(\mathbf{x}) - (G\text{-}1)\tilde{Z}_g(\mathbf{x}), \tag{4.24}$$

where $\tilde{\phi}_g$ represents a weighted adjusted estimate based on how much $\tilde{Z}_g(\mathbf{x})$ deviates from $\tilde{Z}(\mathbf{x})$

$$\tilde{\phi}_g = G \cdot [\tilde{Z}(\mathbf{x}) - \tilde{Z}_g(\mathbf{x})] + \tilde{Z}_g(\mathbf{x}), \tag{4.25}$$

which in turn gives the reduced-bias jackknifed estimator of $Z(\mathbf{x})$

$$\hat{Z}_G(\mathbf{x}) = (G)^{-1} \cdot \sum_{g=1}^{G} \tilde{\phi}_g \tag{4.26}$$

(Lavenberg, Moeller, and Welch 1982; also see Nelson 1990).

Using the jackknifed estimator raises several implementation issues. First, one must decide how to partition I to balance the computation requirements of (4.26) against the need for a lower-bias estimator. Kleijnen (1974) reports that Tocher (1963) shows that for $G = 2$ the variance of (4.26) is twice that for $\hat{Z}_{CV}(\mathbf{x})$ using a non-stochastic $\boldsymbol{b}_v^*$ in (4.22), and therefore suggests $G > 2$ as a technique for reducing it. Although both Kleijnen (1974) and Miller (1974) suggest using $G = I$ and $H = 1$ as the best jackknife grouping, such a configuration would be computationally expensive. As a compromise, this research will assign $H = 10$ and $I \geq 50$. Second, since Σ_v is known from the model's distributional assumptions, one could skip estimating $\hat{\Sigma}_v$, particularly since the assumption of independence of the elements in ω and $\mathbf{T}$ renders the off-diagonal elements in Σ_v zero. However, following Kleijnen (1974) and Nelson (1990), all elements of the sample covariance matrix $\hat{\Sigma}_v$ will be estimated from the sample data in Z using (4.19-20), even though Σ_v is known. (This implies that the off-diagonal elements of $\hat{\Sigma}_v \neq 0$ due to sampling error.) Finally, Lavenberg, Moeller, and Welch (1982) also observe that (4.26) does not require any distributional assumptions; hence, the jackknife procedure will work for any continuous or discrete distributions in ω and $\mathbf{T}$.

4.2.2.3 Non-Stationarity

Adding the k subscript back, a stationary $\boldsymbol{b}_v^*$ is unlikely over all values of $\mathbf{x}$ because of the changes $\mathbf{x}_k$ imposes on (3.1b) through $\mathbf{Tx}$. Therefore, $\hat{\boldsymbol{b}}_v^*$ should be

re-estimated for each $\mathbf{x}_k$ independently from other points during a search technique or experimental design.

4.2.3 Latin Hypercube

4.2.3.1 General Description

Employing Latin Hypercube (LH) sampling as a method for variance reduction was suggested by Wilson (1995) based on his recent research in stochastic activity networks. LH uses an extended stratified sampling structure proven by McKay, Conover, and Beckman (1979) to guarantee that for any estimator of the mean response $\bar{Y}$ associated with the random variable Y, where the equivalent LH estimator is $\bar{Y}_{LH}$, $Y = f(\mathbf{x})$, and x^i (a component of $\mathbf{x}$) represents a random variable, $\text{VAR}[\bar{Y}_{LH}] \leq \text{VAR}[\bar{Y}]$ if Y is a monotonic function for each x^i in $\mathbf{x}$. *Since the objective value changes as a monotonic function of any right-side value of any inequality constraint, it immediately follows that* $\text{VAR}[\hat{Z}_{LH}(\mathbf{x}_k)] \leq \text{VAR}[\hat{Z}_{RS}(\mathbf{x}_k)]$ *for the class of problems represented by (3.1).*

Further motivation for investigating LH variance reduction in the recourse context comes from McKay, Conover, and Beckman:

> One advantage of the Latin hypercube sample appears when the output *Y(t)* is dominated by only a few of the components of **X**. This method insures that each of those components is represented in a fully stratified manner, no matter which components might turn out to be important (McKay, Conover, and Beckman 1979, 240).

This dominance characteristic concurs with the CV assumption of high correlation between one or more elements of $(\omega - \mathbf{Tx})$ and z_{ik}. Second, subsequent empirical

results by Avramidis (1992), Avramidis and Wilson (1995), McKay, Conover, and Beckman (1979), and Stein (1987) found LHs provide considerable variance reduction in a variety of simulation environments. Third, McKay, Conover, and Beckman (1979) show that $E[\hat{Z}_{LH}(\mathbf{x}_k)] = Z(\mathbf{x}_k)$; i.e., the LH estimator is unbiased and thus does not need any compensating adjustments (e.g., the jackknifing procedures of CVs). Finally, unlike CVs, LHs have the advantage of not requiring *a priori* knowledge of a random variable with known correlation to the response. Thus, LHs would be especially useful in the initial search for the region of optimality, as well as during the experimental design phase.

The LH approach falls under the *Correlation Induction* category of VRTs. Like antithetic variates (AVs), LHs attempt to induce a negative correlation between two or more responses such that the variance of the averaged responses is less than that of a randomly sampled estimator. Beginning with the next set of definitions (and dropping the k subscript again for convenience), the accompanying explanation of the LH method for the recourse problem follows the description of the general case provided by Avramidis and Wilson (1995).

Definition. In general, let D_{ID}^{T} represent the joint distribution of T independent variables.

Definition. In general, let the T-variate distribution D_{CI}^{T} possess the following properties:

1. Each univariate marginal distribution of D_{CI}^{T} follows a uniform (0,1) interval.

2. Every bivariate marginal distributions is *negative quadratic dependent* (*NQD*); i.e., for any bivariate random vector (B_1,B_2) in D_{CI}^T then $\text{PROB}\{B_1 \le b_1, B_2 \le b_2\} \le \text{PROB}\{B_1 \le b_1\} \cdot \text{PROB}\{B_2 \le b_2\}$.

3. All bivariate marginal distributions of D_{CI}^T equal each other.

Definition. Let R represent the number of random variables in $\mathbf{T}$ and ω.

Definition. Let U^{tr} represent a random variate drawn on the unit interval (0,1) for the random column vector $\boldsymbol{U}^{\bullet r}$, where $r = 1, \ldots, R$, and $t = 1, \ldots, T$. This implies generating the actual random variables in the recourse problem by using unit interval variates U^{tr} as inputs (e.g., inverse transformations for continuous variables). The elements $U^{\bullet r}$ also represent intervals (or stratum) of equal probability with respect to the original distribution of the random variable represented by the column vector $\boldsymbol{U}^{\bullet r}$.

Definition. Define the $i^{\underline{th}}$ sample matrix $\boldsymbol{U}_i$ to consist of the column vectors $[\boldsymbol{U}_i^{\bullet 1}, \boldsymbol{U}_i^{\bullet 2}, \ldots, \boldsymbol{U}_i^{\bullet R}]$ or equivalently the row vectors $[\boldsymbol{U}_i^{1\bullet}, \boldsymbol{U}_i^{2\bullet}, \ldots, \boldsymbol{U}_i^{T\bullet}]$.

Definition. Let z_{it} represent the value for $\mathbf{cx} + h(\mathbf{x},\omega_{it},\mathbf{T}_{it})$ for the sample row vector of random variates $\boldsymbol{U}_i^{t\bullet}$ from $\boldsymbol{U}_i$.

Definition. Define the function $f_{(3.1)}(\boldsymbol{U}^{t\bullet})$ to represent the value of $\mathbf{cx} + h(\mathbf{x},\omega_{it},\mathbf{T}_{it})$ in (3.1) using the random variables ω - $\mathbf{Tx}$ derived from the input variates $\boldsymbol{U}^{t\bullet}$; i.e., $z_{it} = f_{(3.1)}(\boldsymbol{U}_i^{t\bullet})$.

The general form of a correlation induction sample $\boldsymbol{U}_i$ is

$$\boldsymbol{U}_i = \begin{Bmatrix} U_i^{11} & U_i^{12} & \dots & U_i^{1R} \\ U_i^{21} & U_i^{22} & \dots & U_i^{2R} \\ \vdots & \vdots & \vdots & \vdots \\ U_i^{T1} & U_i^{T2} & \dots & U_i^{TR} \end{Bmatrix} \tag{4.27}$$

where the column vectors $\boldsymbol{U}_i^{\bullet r}$ fall into two mutually exclusive sets — one group following distribution D_{CI}^T while the other conforms to D_{ID}^T (and where the second set of column vectors $\boldsymbol{U}_i^{\bullet r}$ with distribution D_{ID}^T can be empty). This structure guarantees that the estimators z_{it}, $t = 1, \dots, T$ derived from the row vectors $\boldsymbol{U}_i^{t\bullet}$ will be *negatively correlated* since (4.27) provides *dependent* rows $\boldsymbol{U}_i^{t\bullet}$ (hence dependent z_{it}) under the *NQD* property of D_{CI}^T. At the same time, it also insures the column vectors $\boldsymbol{U}_i^{\bullet r}$ *independence* through the random marginal univariate sampling of the unit interval (0,1) (Avramidis and Wilson 1995).

An examination of antithetic variates (AVs) by Avramidis and Wilson (1995) provides an easy example of how (4.27) works. AVs form a special case of (4.27) where $T = 2$; U^{1r}, $r = 1, \dots, R$, are randomly and independently sampled from a uniform distribution (0,1); and, $U^{2r} = (1 - U^{1r})$, $r = 1, \dots, R$. This gives

$$\boldsymbol{U}_i = \begin{Bmatrix} U_i^{11} & U_i^{12} & \dots & U_i^{1R} \\ 1\text{-}U_i^{11} & 1\text{-}U_i^{12} & \dots & 1\text{-}U_i^{1R} \end{Bmatrix} \tag{4.28}$$

where $z_{i1} = f_{(3.1)}(\boldsymbol{U}^{1\bullet})$, $z_{i2} = f_{(3.1)}(\boldsymbol{U}^{2\bullet})$, and dependence exists between $\boldsymbol{U}^{1\bullet}$ and $\boldsymbol{U}^{2\bullet}$ through the relationship $U^{2r} = (1 - U^{1r})$. Furthermore, the random sampling of

U^{1r}, in conjunction with $U^{2r} = (1 - U^{1r})$, meets the criteria of the distribution D^2_{CI}. Hence, the column vectors $\boldsymbol{U}^{\bullet r}$ are both independent and provide a negative correlation between z_{i1} and z_{i2} (Avramidis and Wilson 1995). For sample $\boldsymbol{U}_i$, the antithetic estimate of $Z(\mathbf{x})$ for I sample size is

$$\hat{Z}_{AV}(\mathbf{x}) = (I)^{-1} \cdot \sum_{i=1}^{I} z_i, \tag{4.29}$$

where

$$z_i = \frac{z_{i1} + z_{i2}}{2}, \tag{4.30}$$

and has variance

$$\text{VAR}[\hat{Z}_{AV}(\mathbf{x})] = \frac{\text{VAR}[z_{i1}] + \text{VAR}[z_{i2}] + 2\text{COVAR}[z_{i1};z_{i2}]}{4I}. \tag{4.31}$$

If z_{i1} and z_{i2} were independent $\text{COVAR}[z_{i1};z_{i2}] = 0$ and $\hat{Z}_{AV}(\mathbf{x}) = \hat{Z}_{RS}(\mathbf{x})$; however, since monotonicity guarantees that z_{i1} and z_{i2} are negatively correlated, $\text{COVAR}[z_{i1};z_{i2}] < 0$ and $\text{VAR}[\hat{Z}_{AV}(\mathbf{x})] < \text{VAR}[\hat{Z}_{RS}(\mathbf{x})]$ (Law and Kelton 1991).

Returning to Avramidis and Wilson (1995), LH sampling implements the structure of (4.27) and its supporting assumptions using the relation

$$U^{tr} = \frac{\pi_r(t) - 1 + U^{tr}_{ID}}{T} \tag{4.32}$$

under the following definitions.

Definition. Where the ordering of the set of integers $\{1, \ldots, T\}$ has $T!$ permutations, let the function $\pi_r(\bullet)$, $r = 1, \ldots, R$, represent a random sample (with replacement) of that set of permutations. Furthermore, define $\pi_r(t)$ as the $t^{\underline{th}}$ element in the $r^{\underline{th}}$ random permutation.

Definition. Let U_{ID}^{tr} be an independent random sample of the unit interval (0,1) for $t = 1, \ldots, T$, and $r = 1, \ldots, R$.

Note that the elements of the column vectors $\boldsymbol{U}^{\bullet r}$ represent a uniformly stratified sample of size T randomly ordered by the permutation $\pi_r(\bullet)$. Within each stratum an independent sampling of the unit interval (0,1) is taken, while the permutation function $\pi_r(\bullet)$ insures that every stratum is represented once in each column vector. Consequently, (4.32) describes a distribution D_{CI}^{T}, since each univariate marginal distribution follows a uniform (0,1) interval; and, every bivariate marginal distribution is both *NQD* and equivalent to any other bivariate marginal distribution. Furthermore, since both U_{ID}^{rt} and $\pi_r(\bullet)$ are independent, the column vectors $\boldsymbol{U}^{\bullet r}$ are independent as well. Therefore, by previous definitions (4.32) provides a negative correlation for row vectors $\boldsymbol{U}^{t\bullet}$ by providing one or more independent column vectors $\boldsymbol{U}^{\bullet r}$ with distribution D_{CI}^{T} (Avramidis and Wilson 1995).

4.2.3.2 Discrete Random Variables

Since the problems examined by this research model the stochastic elements of ω and $\mathbf{T}$ with discrete distributions, the LH assumption of continuous variates in (4.32) must be adjusted. McKay (1988) suggests such a modification

by allocating the discrete values of the random variable v^p proportionally to their probabilities within its associated column vector $\boldsymbol{U}^{\bullet r}$.

Definition. Let the random variable v^p take on D discrete realizations w_d, $d = 1, \ldots, D$. Furthermore, let $\boldsymbol{U}_i^{\bullet p}$ denote its associated column vector in $\boldsymbol{U}_i$ for the i^{th} sample v_i^p, and its dimension T represent the stratification size of $\boldsymbol{U}_i$.

Definition. Let $p_d = \text{PROB}\{v^p = w_d\}$, where $\sum_{d=1}^{D} p_d = 1$.

Since the exact allocation of any w_d $(p_d \cdot T)$ will most likely be non-integer, it can be partitioned into its integer and fractional components

$$p_d \cdot T = \text{INT}(p_d \cdot T) + f_d \tag{4.33}$$

such that each w_d of v^p will have at least $\text{INT}(p_d \cdot T)$ elements in $\boldsymbol{U}_i^{\bullet p}$ for each sample i of size T. It follows directly that

$$T = \sum_{d=1}^{D} \text{INT}(p_d \cdot T) + \sum_{d=1}^{D} f_d \tag{4.34}$$

and since both T and $\text{INT}(p_d \cdot T))$ are integer

$$F^p = \sum_{d=1}^{D} f_d \tag{4.35}$$

is also an integer representing the remaining slots in $\boldsymbol{U}_i^{\bullet p}$ that can be filled by additional w_d beyond their guaranteed quota of $\text{INT}(p_d \cdot T)$ (McKay 1988).

McKay's method introduces the column vector independence of $\boldsymbol{U}_i^{\bullet p}$ first by randomly sampling on the unit interval (0,1) F^p times and allocating the additional w_d based on the proportion

$$\frac{f_d}{F^p} \tag{4.36}$$

where

$$\sum_{d=1}^{D} \frac{f_d}{F^p} = 1. \tag{4.37}$$

The resulting selection of size T using $\text{INT}(p_d \cdot T)$ and the random sampling just discussed is then independently ordered within the column vector $\boldsymbol{U}_i^{\bullet p}$ using the permutation function $\pi_p(T)$. This process repeats for each LH sample i, $i = 1, \ldots, I$, by which w_d will appear on average in $\boldsymbol{U}_i^{\bullet p}$ according to the frequency described by (4.33) (McKay 1988). Also note that U_i^{rt} of $\boldsymbol{U}_i$ in the discrete context represents the actual random variable, instead of a random variate drawn on the unit interval requiring (usually) an inverse transformation.

This section completes its description of the LH technique with an example $\boldsymbol{U}_i$ for APL1P. Setting the sample size $T = 10$, TABLE 4.1 gives the $\text{INT}(p_d \cdot T)$, f_d, f_d/F^P, and associated assigned portions of the unit interval for the three different discrete distributions followed by ξ and ω. Figure 4.1 gives a sample $\boldsymbol{U}_i$ based on the parameters in TABLE 4.1; a random selection of two additional realizations for ω^j, $j = 1, 2, 3$, which are (1100,1000), (1100,1000), and (1100,1200), respectively; and, the random permutation function $\pi(T)$.

TABLE 4.1
LH/MCKAY DISTRIBUTION OF APL1P STOCHASTIC PARAMETERS FOR $T = 10$

v^p	w_d	PROB$\{v^p = w_d\}$	INT$(p_d \cdot T)$	f_d	$\frac{f_d}{F^P}$	Proportion of (0,1) Itvl.
	1.0	.2	2	.0	.0	—
ξ^1	.9	.3	3	.0	.0	—
	.5	.4	4	.0	.0	—
	.1	.1	1	.0	.0	—
	1.0	.1	1	.0	.0	—
	.9	.2	2	.0	.0	—
ξ^2	.7	.5	5	.0	.0	—
	.1	.1	1	.0	.0	—
	.0	.1	1	.0	.0	—
	900	.15	1	.5	.25	.0 - .25
ω^j	1000	.45	4	.5	.25	$.25^+$ - .5
	1100	.25	2	.5	.25	$.5^+$ - .75
	1200	.15	1	.5	.25	$.75^+$ - 1.0

	$U^{\bullet\xi 1}$	$U^{\bullet\xi 2}$	$U^{\bullet\omega 1}$	$U^{\bullet\omega 2}$	$U^{\bullet\omega 3}$
$U^{0\bullet}$	.5	.7	1000	900	900
$U^{1\bullet}$	1.0	.1	1000	1100	1200
$U^{2\bullet}$	.5	.7	900	1000	1100
$U^{3\bullet}$	.9	1.0	1100	1000	1000
$U^{4\bullet}$	1.0	.9	1000	1000	1000
$U^{5\bullet}$	.5	.7	1000	1100	1000
$U^{6\bullet}$	.1	.9	1100	1200	1200
$U^{7\bullet}$	.5	.0	1100	1100	1100
$U^{8\bullet}$	.9	.7	1200	1000	1100
$U^{9\bullet}$	.9	.7	1000	1000	1000

Figure 4.1. Sample U_i LH Matrix ($T = 10$)

4.2.4 Example Variance Reduction Results

Table 4.2 and Figure 4.2 provide variance reduction results for APL1P using CV, LH, and random sampling (RS) methods. Since the search techniques and response surface procedure use a *single* estimate $\hat{Z}(\mathbf{x}_k)$ for each $\mathbf{x}_k$, this example compares the accuracy of the sample estimators and magnitude of the sample variance under the following definitions.

> **Definition.** Where $\hat{Z}_s(\mathbf{x}_k)$ represents an unbiased estimator for $Z(\mathbf{x}_k)$ as previously defined for sample size I using sampling technique $s \in \{RS,CV,LH\}$, let $\hat{Z}_s(\mathbf{x}_k)_n$ be the n^{th} unbiased estimate independently sampled from any other estimate of $Z(\mathbf{x}_k)$.

Since $\hat{Z}_s(\mathbf{x}_k)_n$ itself is a random variable let

$$\bar{Z}_s(\mathbf{x}_k)_N = (N)^{-1} \cdot \sum_{n=1}^{N} \hat{Z}_s(\mathbf{x}_k)_n \tag{4.38}$$

represent the mean for N independent sample estimates of $Z(\mathbf{x}_k)$ using sampling technique s, while

$$\bar{s}^2{}_s(\mathbf{x}_k)_N = (N\text{-}1)^{-1} \cdot \sum_{n=1}^{N} (\hat{Z}_s(\mathbf{x}_k)_n - \bar{Z}_s(\mathbf{x}_k)_N)^2 \tag{4.39}$$

defines the sample variance of the estimator $\bar{Z}_s(\mathbf{x}_k)_N$. These definitions parallel the previous formulations (4.1) and (4.3) for a single estimator $\hat{Z}_s(\mathbf{x}_k)$ of $Z(\mathbf{x}_k)$, except that here the variance of interest is the one associated with the estimator itself. Justification for such a comparison follows directly from the fact that as an

unbiased estimate of $Z(\mathbf{x}_k)$, any $\hat{Z}_s(\mathbf{x}_k)$ drawn from the sampling technique with the smallest $\bar{s}^2_s(\mathbf{x}_k)$ will — *on average* — be closer to $Z(\mathbf{x}_k)$ than $\hat{Z}_t(\mathbf{x}_k)$, $t \neq s$ (see Law and Kelton 1991).

For comparison purposes in APL1P the random sample and control variate estimators $\hat{Z}_{RS}(\mathbf{x}_k)_n$ and $\hat{Z}_{CV}(\mathbf{x}_k)_n$ use a sample size $I = 50$. The CVs use ξ^1, ξ^2, and ω^1 as the control variates with a jackknife partition $G = 5$ and $H = 10$. Lacking a preliminary screening design, these controls are selected based on the direct influence of ξ^1 and ξ^2 on the amount of resource x^1 and x^2, while adding ω^1 due to its association with the highest recourse cost coefficients. (Adjusting previous notation, let $\hat{Z}_{CV}(\mathbf{x}_k)_n = \hat{Z}_G(\mathbf{x}_k)_n$; i.e., the CV designation assumes the jackknifed adjustment.) The Latin Hypercube estimator $\hat{Z}_{LH}(\mathbf{x}_k)_n$ uses a stratification size $T = 50$ as well, equivalently giving it the same sample size as RS and CV. For vectors $\mathbf{x}_k$, $k = 1,\ldots, 5$, ten independent estimators $\hat{Z}_s(\mathbf{x}_k)_n$ ($s \in$ {RS,CV,LH}, $N = 10$) provide the basis for calculating $\bar{Z}_s(\mathbf{x}_k)_N$ and $\bar{s}^2_s(\mathbf{x}_k)_N$.

The results in Table 4.2 show a dramatic reduction in the variance of the estimator for both CVs and LHs, although in general LHs gave both the lowest variance and the most accurate estimate of $Z(\mathbf{x}_k)$ (with the exception of CVs for $\mathbf{x}$ = (2700,2700)). Similarly, Figure 4.2 graphically demonstrates the LH (and to a lesser extent the CV) estimates varying far less than their RS counterparts about a more accurate assessment of $Z(\mathbf{x})$ by plotting the individual $\hat{Z}_s(\mathbf{x}_k)_n$ used to calculate Table 4.2 against $Z(\mathbf{x}_k)$ and their respective $\bar{Z}_s(\mathbf{x}_k)_N$. Indeed, these plots can be thought of as sample distributions of the various estimation techniques $s \in$ {RS,CV,LH}, for $N = 10$, and $\hat{Z}_s(\mathbf{x}_k)$ as a single sample. From inspecting Figure

TABLE 4.2
COMPARISON OF ESTIMATOR ACCURACY AND VARIANCE FOR RS, CV, AND LH SAMPLING TECHNIQUES FOR APL1P(I=50, N=10)

$\mathbf{x}_k$	Z(x)	$\bar{Z}_{RS}$	$\bar{s}^2_{RS}$*	$\bar{Z}_{CV}$	$\bar{s}^2_{CV}$*	%†	$\bar{Z}_{LH}$	$\bar{s}^2_{LH}$*	%†
(1800,1800)	24689	25065	623	24875	44	.93	24677	15	.98
(900,900)	26425	26989	229	26795	49	.79	26450	6	.98
(900,2700)	25131	25646	856	25424	38	.96	25123	6	.99
(2700,900)	25299	25806	826	25395	51	.94	25309	12	.99
(2700,2700)	27499	27576	342	27552	35	.90	27618	61	.82

* - in thousands † - % variance reduction from RS

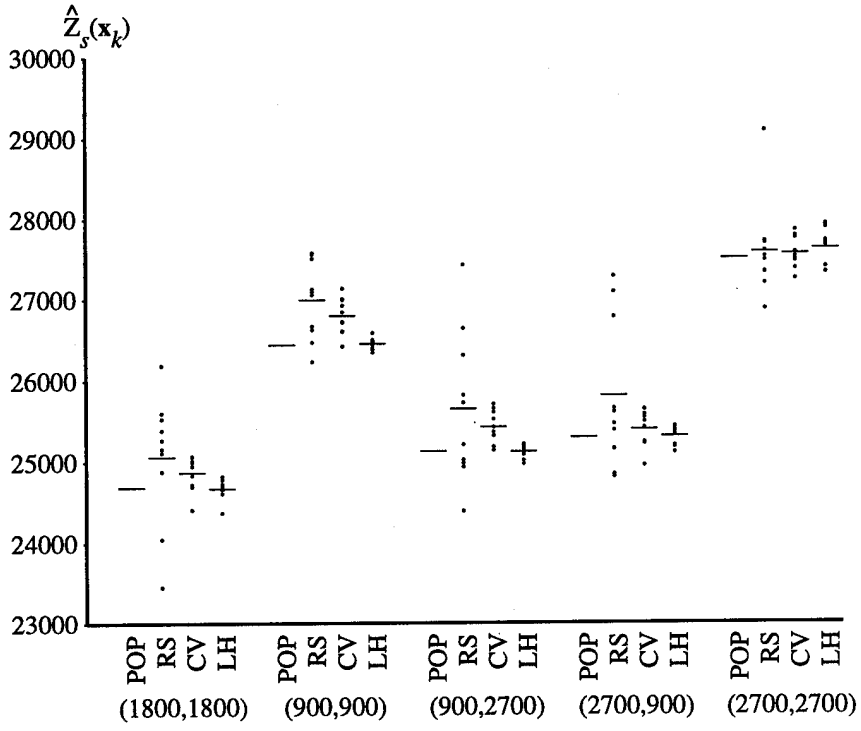


Figure 4.2. Scatterplot of $\hat{Z}_s(\mathbf{x}_k)_n$ [·] and $\bar{Z}_s(\mathbf{x}_k)_N$ [–] for Population, RS, CV, and LH Sampling Techniques

4.2 and knowing the true mean $Z(\mathbf{x}_k)$, LH sampling clearly stands out as the most accurate method.

Figure 4.2 also gives the most powerful evidence of how the variability of $\hat{Z}_{RS}(\mathbf{x}_k)_n$ can badly mislead either the search or the response surface approximation. For instance, examining a near optimal solution $\mathbf{x} = (1800,1800)$ finds one extreme of $\hat{Z}_{RS}(\mathbf{x}_k) = 26186$ while the other extreme is 23449 (the true population mean, $Z(\mathbf{x} = (1800,1800)) = 24689$, is designated POP). Most importantly, *both results are far more likely to occur as* $\hat{Z}_{RS}(\mathbf{x}_k)$ *samples than either* $\hat{Z}_{CV}(\mathbf{x}_k)$ *or* $\hat{Z}_{LH}(\mathbf{x}_k)$ *for estimates of* $Z(\mathbf{x})$. Furthermore, the consequences of such error becomes obvious given $Z(\mathbf{x}^*) = 24642$ and the 'flatness' of the region of optimality. If either estimate had been used in a search technique, it would have misdirected the process in a region sensitive to sampling errors. Similarly, in the context of a response surface design, had the lower estimate constituted a non-central design point the regression would assume the objective function value for (1800,1800) was *lower* than the presumed optimal centerpoint and fitted the appropriate — *but incorrect* — non-convex saddle surface.

4.3 Experimental Design

4.3.1 Introduction

Once the search process has found one or more optimal (or near-optimal) solutions, the next step requires deriving a polynomial approximation of the response $Z(\mathbf{x})$ to changes in $\mathbf{x}$. Since no two problems will exhibit the same response characteristics, this dissertation cannot give precise guidance on issues associated with constructing the design (i.e., factor selection, factor ranges, and

type of design) used to calculate the response approximation; nor can it turn to the stochastic programming literature for help since this idea has not been tried before in the context of linear programming with recourse. Therefore, this section will present a general approach for constructing designs as applied to (3.1a) based on the experimental design and simulation literature, and leave the details to the specifics of the problem.

Instead of constructing a design to derive the response surface, one possibility would be to simply estimate a regression model based on the data collected by the search process. However, such an unstructured procedure would not provide as good an approximation as a more formal experimental design for several reasons. Quoting Law and Kelton

> In simulation, *experimental design* provides a way of deciding before the runs are made which particular configurations to simulate so that the desired information can be obtained with the least amount of simulating. Carefully designed experiments are much more efficient than a "hit-or-miss" sequence of runs in which we simply try a number of alternative configurations unsystematically to see what happens (Law and Kelton 1991, 657).

A better approach would treat the search data as a type of reconnaissance information that gives us much-needed insight into the basic behavior of the recourse problem and provides guidance on how to structure the formal experimental design. Specifically, this preliminary information directs the centering of the design based upon the location of the optimal solution; and, through its rudimentary knowledge of the relationship between $\mathbf{x}$ and $Z(\mathbf{x})$, restricts the design (and associated response surface) to the region of optimality. By using this two-step method (preliminary screening and experimental design)

one can easily assess the accuracy of the resulting polynomial approximation using the characteristics of the specific type of design in conjunction with the known convexity of $Z(\mathbf{x})$.

Although the topic of experimental design encompasses an immense literature (Steinberg and Hunter 1984), this research focuses on its application within the realm of simulation. Law and Kelton (1991) cite three specific advantages simulation experiments have over their physical or industrial counterparts:

1. *Control.* Unlike a physical environment, simulation experiments control the level or value of the input factors. In the recourse context, this means the first-stage decision vector $\mathbf{x}$.

2. *Variability.* Controlling the source of variability — the random number generators — allows for the application of the VRTs explained in the previous section.

3. *Randomization.* Unlike physical environments, where systematic error requires repeated experiments under the same input values (termed *replications*), a single estimate of sufficient sample size for each set of input factors will suffice (Law and Kelton 1991).

These advantages in turn lead to designs whose estimates of the polynomial parameters are both relatively simple to calculate and highly efficient (Box and Draper 1987).

This section will introduce the experimental designs used for this research by first reviewing its terminology and assumptions, then presenting the type of designs used to derive the response surfaces, and concluding with a simple example for APL1P. Topics dealing specifically with response surface estimation are deferred until Section 4.4.

4.3.2 Terminology and Assumptions

This dissertation uses the following terminology from the experimental design literature as applied in the context of the recourse problem.

Definition. Let k index a partition of size K of the decision vector $\mathbf{x} \in X$ such that $\mathbf{x} = [x^1, \ldots, x^k, \ldots, x^K \mid x^{K+1}, \ldots, x^{n(3.1a)}]^T = [\mathbf{x}_K \mid \mathbf{x}_{K'}]^T$ (note that k no longer refers to a vector $\mathbf{x}_k$ in the context of the search sequence).

Definition. Let the term *factor* be synonymous with x^k, $k = 1, \ldots, K$ in $\mathbf{x}_K$. The subset of factors $\mathbf{x}_K$ constitutes an assumption of which variables in $\mathbf{x}$ significantly influence $Z(\mathbf{x})$. Since the non-factors $x^{K+1}, \ldots, x^{n(3.1a)}$ in $\mathbf{x}_{K'}$ are held constant, the experimental design consists solely of changes in the *factor* values as inputs, with $\hat{Z}_s(\mathbf{x})$ as the output (or response).

Definition. The literature refers to *levels* as the values of the factors, which in (3.1a) corresponds to the value for $x^k \geq 0$. *N-level* designs refer to the number of N levels each factor is restricted to having in the experimental design. This research uses only two-level designs (disregarding center point and axial values), and in keeping with

conventional notation designates the lowest value for x^k as '-' and the highest as '+'.

Definition. *Center point* refers to the intermediate or average value of x^k, and is represented as '0'; by extension, the centerpoint of the experimental design for $\mathbf{x}_K$ is '**0**'.

Definition. *Design point* refers to a specific combination of factor levels for $\mathbf{x}_K$.

Definition. *Axial point* denotes a design point whose factor levels are all center points '0' except for $x^k = \pm\alpha^k$, where α^k is greater than the value for x^k represented by '+' (and by symmetry $-\alpha^k <$ '-').

Definition. *Response* refers to the output of the simulation — in this case the estimator $\hat{Z}_s(\mathbf{x})$ for $Z(\mathbf{x})$. The *response surface approximation* $V(\mathbf{x})$ refers to a second-order polynomial approximation

$$V(\mathbf{x}_K) = \hat{a}_0 + \sum_{k=1}^{K} \hat{a}_k x^k + \sum_{i \geq j}^{K} \sum_{j=1}^{K} \hat{a}_{ij} x^i x^j \qquad (4.40)$$

for a specified range of $\mathbf{x}_K$, where $Z(\mathbf{x}) = V(\mathbf{x}_K) + \varepsilon(\mathbf{x}_K)$, and $\varepsilon(\mathbf{x}_K)$ is error due to random variability and lack of fit.

Definition. *Replication* refers to a single simulation and its associated response at a given design point. In the recourse context, the estimator $\hat{Z}_s(\mathbf{x})$ for the sampling technique $s \in \{RS,CV,LH\}$ represents a single replication for the design point defined by the factor levels of $\mathbf{x}_K$.

Definition. Letting r^k represent the absolute value of the difference between the high value '+' and low value '-' for x^k, and c^k the value of the center point for x^k, the *coded* value x_c^k for x^k is

$$x_c^k = \frac{x^k - c^k}{r^k} \quad . \tag{4.41}$$

By this definition, the coded values for the two-level design points are +1 and -1, the center points 0, and the axial points $\alpha^k > 1$, $-\alpha^k < -1$.

With these basic definitions, the literature consists of a wide variety of experimental design structures (Central Composite, Box-Behnken, Mixture, Plackett-Burman, to name a few) that provide their associated response surface approximations with different degrees of higher order interactions, confounding, coefficient variance, response variance, and orthogonality. This dissertation restricts its investigation to a widely used design — the *central composite* (CC) design — as the principal method (with variations) for deriving the polynomial approximations of Z(**x**) (Box, Hunter, and Hunter 1978, Diamond 1989, Kleijnen 1987, Law and Kelton 1991, Lorenzen and Anderson 1993, or Montgomery 1984).

4.3.3 Central Composite Design

Following Montgomery (1984), the CC design represents a very popular design for fitting second-order models, and contains three basic components — a 'core' 2^K factorial or fractional design, $2K$ axial points and n_K centerpoints.

1. *Factorial Design.* The 2^K *factorial* design consists of all combinations of factor levels for a two-level design; graphically, it would represent the four corner points of a square for 2 variables, the eight corner points of a cube for 3 variables, etc. A 2^{K-P} *fractional* design consists of a subset of the full factorial design (minus 2^P design points), and is referred to as a $1/2^P$ fractional design (discussed in more detail below).

2. *Axial Points.* Continuing with the geometric analogy, the axial points represent design points projected through the center of each side of the hypercube. Mathematically stated, the sets of axial points would be $\{\pm\alpha^1, 0, \ldots, 0\}$, $\{0, \pm\alpha^2, 0, \ldots, 0\}$, $\ldots$, $\{0, 0, \ldots, \pm\alpha^K\}$, and for a CC design of K factors constitutes $2K$ design points.

3. *Center Points.* The center points represent n_K replications of the design point $\{0, 0, \ldots, 0\}$.

This configuration provides the three minimum levels for each factor necessary to estimate a quadratic model, and depending on the selection of the values for α^k and n_K gives the response approximation the following characteristics.

1. *Rotatable.* The variance of the predicted response $V(\mathbf{x}_K)$ varies only with distance — and not the direction — from the centerpoint. The value of α determines this attribute of the design.

2. *Orthogonal.* This condition occurs whenever the off-diagonal elements of $\mathbf{X}^T\mathbf{X}$ are zero, and is a function of the value of n_K. In the CC design context it minimizes the variance of the regression coefficients $\hat{a}_{ij}$.

3. *Uniform Precision.* Under this property, controlled by the size of n_K, the variance of the response $V(\mathbf{x}_K)$ is consistent throughout the hypercube defined by the factorial design points. Note that the value of n_K for a uniform precision design can differ from that for an orthogonal one.

Another often cited characteristic of CC designs observes that they can be built up from simpler first-order designs that use just the factorial portion (Montgomery 1984). However, given the known convexity of $Z(\mathbf{x})$ this research assumes that a positive definite quadratic approximation will always be required in the region of optimality, and therefore will proceed directly with some version of a CC design.

Resolution, another important concept associated with experimental designs, measures the level of factor interaction the polynomial approximation can independently determine. Recalling the definition of the 2^{K-P} fractional as a subset of the full factorial design minus 2^P design points, the design resolution denotes the level of confounding between certain factor interactions that occurs in fractional designs. In the context of problem (3.1), for example, this means that if both interactions $x^g x^h$ and $x^i x^j x^k$ significantly affect $Z(\mathbf{x})$, certain fractional designs could not discern one effect from the other.

Definition. Define the resolution level R (denoted using Roman numerals) as one where every r-factor interaction can be independently

determined from any other factor interaction containing less than R - r factors (Box, Hunter, and Hunter (1978).

Using the previous example, as two-level and three-level interactions, respectively, $x^g x^h$ and $x^i x^j x^k$ confounding implies no greater than a resolution V fractional design since the difference of the design level (5) and interaction level of the first factor set (2) is not less than the level of the second factor set (3). By contrast, for a resolution V design main factor effects are not confounded with two-way interactions (5 - 1 = 4, which is greater than 2), nor would two-way interactions be confounded with other two-way interactions (5 - 2 = 3, which is still greater than 2). Note that only the full 2^K factorial design provides complete factor interaction estimates. (Also see Box and Hunter 1961.)

One problem associated with the CC design stems from the exponential growth in the factorial portion that occurs with an increase in the number of input factors K. For smaller problems (such as APL1P, CEP1, PGP2 and even 4TERM) such growth is manageable; however, for larger problems (e.g. 20TERM) a full factorial CC design can not be accomplished and analyzed in a reasonable period of time. For example, 20TERM requires $2^{63} \approx 9.223 \cdot 10^{18}$ runs just for the full factorial portion of the design. Consequently, this research employs two basic tactics to reduce the CC design to a manageable level for the larger problems — *preliminary screening* and *fractional factorial* designs.

4.3.3.1 Preliminary Screening

As explained by Box and Draper (1987) the orthogonal (or near-orthogonal) property available through CC designs provides far more precise estimates of $\hat{a}_{ij}$ that give smaller, but more inclusive, individual and joint confidence regions of the regression coefficients. For this reason alone a regression based on the data collected from the search process would not provide as accurate a polynomial approximation. However, the optimal search process can provide preliminary information to assist the development of the CC design in varying degrees, depending upon the size of (3.1).

First, for *all* problems the search technique provides an optimal solution that determines the center point of the design. Centering the design around an optimal solution gives two principal advantages:

1. *Region of Optimality*. The purpose of the response surface analysis is to find a second-order polynomial approximation of the optimum neighborhood, which will often be just a small portion of $\boldsymbol{X}$. Centering the design guarantees that the regression approximates the correct area of interest.

2. *Convexity*. The literature gives this analysis the advantage of knowing the true surface of $Z(\mathbf{x})$ to be convex, which implies that the fitted polynomial should also be convex. Therefore, centering the design on a 'side' of the surface (instead of the 'bottom') — in conjunction with $\hat{Z}_s(\mathbf{x})$ error, the 'flatness' and shape of $Z(\mathbf{x})$, and the design resolution — increases the risk

of the regression fitting a saddle surface. (Box and Draper (1987) give an excellent discussion of this phenomenon in Chapter 11.)

For these reasons, design location is very crucial; therefore, analysis of the search data complements the 'traditional', more formal screening designs used in the experimental literature (e.g., Plackett and Burman 1946). Such designs would not be appropriate without some knowledge of the optimal solution; and, in any case would be difficult to construct given the size of $\boldsymbol{X}$.

Another reason for using the optimal search data in a preliminary analysis concerns the range of the experimental design factors $\mathbf{x}_K$. Since the purpose of the response analysis is to approximate the region of optimality, the design does not need to stretch over the entire set $\boldsymbol{X}$. (Indeed, such a fit would strain the assumption of a quadratic approximation for $Z(\mathbf{x})$.) For smaller problems, determining the range (i.e., the actual values for x^k denoted by '-' and '+') presents no challenge. However, bigger problems will most likely require a formal screening design process.

Highly fractionated designs of resolution III and IV present one method for determining the composition and factor ranges of $\mathbf{x}_K$ using data from the optimal search process. According to Box, Hunter, and Hunter (1978) the Plackett-Burman designs (1946) offer a way to determine the main effects of K factors in $N = K + 1$ runs, where N is a multiple of 4. Similarly, for N as a power of 2, resolution III designs can also be built for $K = N - 1$ factors in N runs by saturating a 2^d factorial design, where $d = \ln(N)/\ln(2)$. If a resolution IV design is needed (i.e., main effects not aliased with two-way interactions), then a *foldover*

of a resolution III design (where the resolution III design is replicated with the '-' and '+' design points reversed and then added to the original design) gives such a design for $2N$ runs (Box, Hunter, and Hunter 1978). Assuming $Z(\mathbf{x})$ contains no complicated higher-order interactions, resolution IV designs should adequately characterize $\mathbf{x}_K$. Finally, note that these designs only estimate linear effects. They are equivalent to the 'core' factorial portion of a CC design, and thus do not include axial or center points.

Finally, *group screening* offers another approach for removing non-important factors before building a CC design. Kleijnen (1987) describes this method where individual factors are consolidated into groups and subsequently treated as a single factor in the screening design. If any ensuing group effect is inconsequential, then all the factors in such a group are assumed insignificant.

4.3.3.2 Fractional Factorial Designs

Once the subset of influential factors $\mathbf{x}_K$ has been set, the next step requires estimating the second-order polynomial equation. Assuming that no third-order or higher interactions occur (which is equivalent to the fundamental assumption this thesis makes regarding the adequacy of a quadratic approximation), fractional CC designs provide another obvious method for reducing the size of the experiment. However, adequately estimating a quadratic function with the assurance of no confounding of the two-way interactions with each other requires at least a resolution V design (Box and Draper 1987). Whitwell and Morbey (1961) note that resolution V designs require a minimum $[1 + K\cdot(K + 1)/2]$ factorial design; when combined with $2K$ axial points and n_K

center points, such designs should handle up to $K = 25$ first stage variables. Assuming that a reasonable reduction from the original **x** occurs in the preliminary screening phase, resolution V designs will be adequate for most problems.

For situations where $K > 25$, resolution III* designs provide another way of estimating a quadratic polynomial. As described by Draper and Lin (1990), resolution III* designs essentially use the axial points of a CC design to de-alias the confounding effects between the one-way and two-way interactions of the factorial portion (also see Hartley 1959). Their article also shows how resolution III* designs can be derived from resolution V designs.

4.3.4 Example Experimental Design

As a small problem, APL1P does not require either a preliminary screening design or a fractional factorial design. Table 4.3 gives the CC design for APL1P, while Figure 4.3 shows the design superimposed on the contour chart for APL1P. Table 4.4 gives the regression results derived from the SAS analysis for both coded and uncoded variable inputs. Finally, because these results represent the population means (hence no variance) the design includes only one center point sample. (Note that this experimental design provides the contour representation of Z(**x**) presented throughout Chapters 3 and 4 (e.g., Figure 4.3). Although technically incorrect since several of them exceed the design region, the contours are nonetheless used for illustrative purposes.)

TABLE 4.3
CENTRAL COMPOSITE DESIGN FOR APL1P

Uncoded x_K		*Coded x_K*		*Response*
x^1	x^2	x^1	x^2	Z(x)
1000	770	-1	-1	26580.3
1000	2370	-1	+1	24862.3
2600	770	+1	-1	25286.7
2600	2370	+1	+1	26622.1
669	1570	-1.414	0	25489.3
2931	1570	+1.414	0	26166.5
1800	439	0	-1.414	26183.3
1800	2701	0	+1.414	25760.9
1800	1570	0	0	24642.6

TABLE 4.4
REGRESSION RESULTS FOR CC DESIGN IN TABLE 4.3

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	R Square
Model	5	4061687	812337	.9894
Error	3	43673	14558	
Total	8	4105360		

Parameter Estimates

Variable	Coded Par. Est.	Uncod. Par. Est.
Intercept	24643.0	33276.0
x^1	180.0	-4.8969
x^2	-122.5	-5.4860
$x^1 \cdot x^2$	763.3	0.0012
$(x^1)^2$	577.1	0.0009
$(x^2)^2$	649.2	0.0010

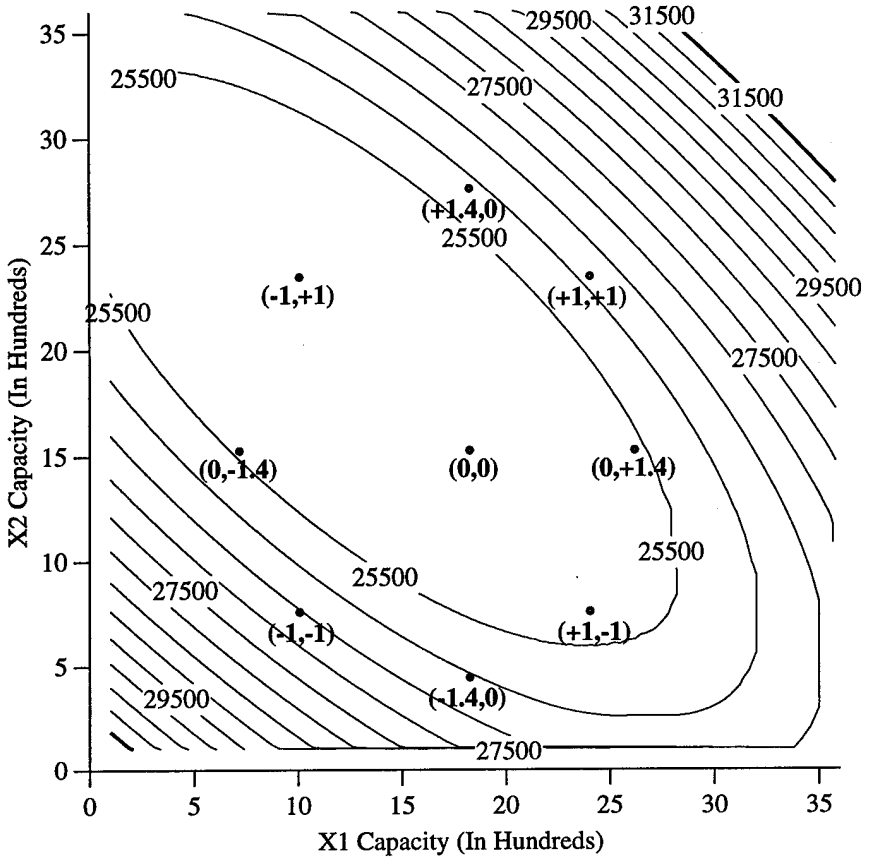


Figure 4.3. Central Composite Experimental Design for APL1P (Coded Variable Designation)

4.4 RESPONSE SURFACE ANALYSIS

4.4.1 Introduction

The application of response surface methodology (RSM) to the two-stage stochastic linear programming problem with recourse represents one of this dissertation's major contributions to this area of research. Specifically, this thesis adopts Box and Draper's (1987) advocacy of examining the type and nature of

factor dependence, a term they use to describe a response function characteristic where its reaction to one factor is not independent of the other factor levels. Such factor interaction typically generates a *ridge system* of responses that can take on various shapes and levels of stationarity, symmetry, and attenuation. This type of response analysis provides the best insight on the reaction of the response to changes in the input variables for the following reasons:

1. *Alternative Optima.* A range of alternative optimal solutions often can be found along a maxima or minima ridge, where changes in one input variable can be compensated by changes in another with no loss of optimality. The direction of the ridge measures this exchange ratio, and therefore may find more suitable solutions in practical or subjective terms.

2. *Optimization of Second Response.* Superimposing a second response on the original plot allows for selecting a point along the ridge that optimizes the second response.

3. *Direction of Insensitivity.* Essentially an extension of item (1), the ridge can also give an attenuation direction that minimizes departures from the optimal solution.

4. *Yield Improvement.* For *rising* ridge systems, this analysis gives an improving direction.

5. *Underlying Mechanism.* Factor dependence can supplement the analyst's or decision-maker's knowledge of the underlying mechanisms of the problem (Box and Draper 1987).

Items (1), (3), and (5) promise to be the most advantageous in the context of the recourse problem (3.1). Secondary response surfaces in item (2) are not investigated, and item (4) should not occur if the experimental design is properly centered. Additionally, this research will also emphasize the opposite aspect of item (3); i.e., where *not* to go by finding the *direction of maximum sensitivity.*

Obviously, for higher-dimensional problems such analysis requires an algebraic description of the ridge. Box and Draper (1987) describe the technique of *canonical analysis* (CA) as one method of providing such a description. (In an appendix they also review an alternative analytical method referred to as *ridge analysis* by A. Hoerl (1959) and R. Hoerl (1985), which essentially functions as a steepest ascent technique for second-order surfaces. This dissertation does not employ this type of analysis, and its use of the term 'ridge analysis' refers to identifying the minima and maxima ridge by way of canonical analysis of the response surface advocated by Box and Draper (1987)). The outline of the CA method in the following sections follows the one provided by Box and Draper (1987) based on the two models they present — the *'A'* and *'B'* canonical forms. Additionally, this study demonstrates the application of these methods to APL1P.

4.4.2 *A* Canonical Form

Adapting Box and Draper's (1987) presentation to the recourse problem, describing the coefficients of (4.40) in the matrix forms

$$\mathbf{x}_K = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^K \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_K \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \hat{a}_{11} & 1/2\hat{a}_{12} & \cdots & 1/2\hat{a}_{1K} \\ 1/2\hat{a}_{12} & \hat{a}_{22} & \cdots & 1/2\hat{a}_{2K} \\ \vdots & \vdots & \cdots & \vdots \\ 1/2\hat{a}_{1K} & 1/2\hat{a}_{2K} & \cdots & \hat{a}_{KK} \end{bmatrix} \tag{4.42}$$

allows the fitted second-order response surface approximation of $Z(\mathbf{x})$ to be

$$V(\mathbf{x}_K) = \hat{a}_0 + \mathbf{x}_K^T\mathbf{a} + \mathbf{x}_K^T\mathbf{A}\mathbf{x}_K. \tag{4.43}$$

(Recall that the experimental design uses a subset of first-stage decision variables $\mathbf{x}_K$ of size K, where $\mathbf{x} = [\mathbf{x}_K \mid \mathbf{x}_{K'}]^T \in \boldsymbol{X}$.) Letting λ^k and $\mathbf{m}^k$ represent the eigenvalue and eigenvector, respectively, for $\mathbf{A}$ for $k = 1, \ldots, K$, then

$$\mathbf{A}\mathbf{m}^k = \mathbf{m}^k\lambda^k. \tag{4.44}$$

Where each eigenvector is normalized (i.e., $(\mathbf{m}^k)^T \cdot \mathbf{m}^k = 1$), the matrix $\mathbf{M}$ consists of $\mathbf{m}^k$, $k = 1, \ldots, K$ as its column vectors. Letting Λ be a diagonal matrix whose elements Λ^{kk} are λ^k, $k = 1, \ldots, K$, gives

$$\mathbf{A}\mathbf{M} = \mathbf{M}\Lambda. \tag{4.45}$$

Since $\mathbf{M}$ is an orthonormal matrix $\mathbf{M}^T = \mathbf{M}^{-1}$ and $\mathbf{M}\mathbf{M}^T = \mathbf{I}$; consequently, multiplying (4.45) with $\mathbf{M}^T$ produces

$$\mathbf{M}^{\mathrm{T}}\mathbf{A}\mathbf{M} = \Lambda, \tag{4.46}$$

and, using the second identity with the associative property, rewrites (4.43) as

$$V(\mathbf{x}_K) = \hat{a}_0 + (\mathbf{x}_K^{\mathrm{T}}\mathbf{M})(\mathbf{M}^{\mathrm{T}}\mathbf{a}) + (\mathbf{x}_K^{\mathrm{T}}\mathbf{M})\mathbf{M}^{\mathrm{T}}\mathbf{A}\mathbf{M}(\mathbf{M}^{\mathrm{T}}\mathbf{x}_K). \tag{4.47}$$

Defining $\mathbf{X} = \mathbf{M}^{\mathrm{T}}\mathbf{x}_K$ and $\theta = \mathbf{M}^{\mathrm{T}}\mathbf{a}$ restates (4.47) as

$$V_A(\mathbf{X}) = \hat{a}_0 + \mathbf{X}^{\mathrm{T}}\theta + \mathbf{X}^{\mathrm{T}}\Lambda\mathbf{X} \tag{4.48}$$

or equivalently

$$V_A(\mathbf{X}) = \hat{a}_0 + \sum_{k=1}^{K} \theta^k X^k + \sum_{k=1}^{K} \lambda^k (X^k)^2 \tag{4.49}$$

where $\theta = [\theta^1, \ldots, \theta^K]^{\mathrm{T}}$, $\mathbf{X} = [X^1, \ldots, X^K]^{\mathrm{T}}$, $V_A(\mathbf{X})$ represents the response approximation for the transformed vector $\mathbf{X}$, and by previous definitions $V_A(\mathbf{M}^{\mathrm{T}}\mathbf{x}_K) = V(\mathbf{x}_K)$. Thus, the linear transformation (4.48) forms the *A* canonical configuration where essentially an axis rotation eliminates the cross-product terms in the original response approximation. Furthermore, as explained in more detail shortly, the eigenvalues λ^k indicate the type of surface (4.40) fits, the eigenvectors $\mathbf{M}^k$ denote the component contribution of the original axes to the rotated ones, and θ^k measures the slope of the rotated axes from the origin of the original coordinate system. Finally, setting the derivative of (4.49) with respect to X^k to zero produces

$$X_s^k = \frac{-\theta^k}{2\lambda^k} \tag{4.50}$$

which defines the stationary (minimum) point with respect to the rotated axes (Box and Draper 1987).

Using the coded parameter estimates from Table 4.4, an *A* canonical analysis for APL1P produces

$$\mathbf{a} = \begin{bmatrix} 179.994 \\ -122.495 \end{bmatrix}, \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} 577.082 & 381.670 \\ 381.670 & 649.194 \end{bmatrix}, \tag{4.51}$$

which in turn gives

$$\mathbf{M} = \begin{bmatrix} .7396 & .6370 \\ -.6370 & .7396 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 229.769 & 0.000 \\ 0.000 & 996.507 \end{bmatrix}, \quad \theta = \begin{bmatrix} 214.089 \\ 29.195 \end{bmatrix} \tag{4.52}$$

the linear transformation

$$V_A(\mathbf{X}) = 24643 + 214.09X^1 + 29.19X^2 + 229.8(X^1)^2 + 996.5(X^2)^2, \tag{4.53}$$

and the stationary points in the rotated coordinates as

$$X_s^1 = \frac{-214.09}{2 \cdot 229.8} = -.4658 \quad \text{and} \quad X_s^2 = \frac{-29.19}{2 \cdot 996.5} = -.0146\,. \tag{4.54}$$

Figure 4.4 illustrates the *A* canonical analysis of APL1P. Since the coded variables' axes (not shown) center around the optimal solution $\mathbf{x}_0 = (1800,1570)$, so do the rotated axes X^1 and X^2 indicated by the solid lines. The axes $\tilde{X}^1$ and $\tilde{X}^2$ displayed by the dashed lines and centered about the stationary point $\mathbf{X}_s$ concern the *B* canonical analysis discussed shortly.

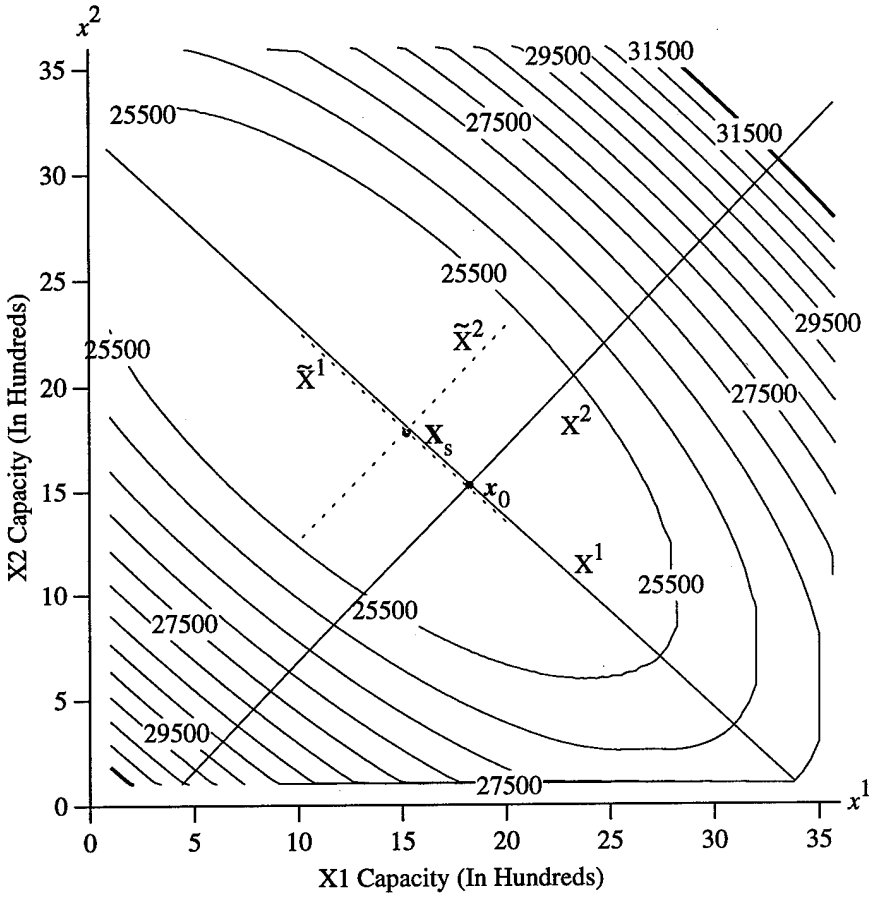


Figure 4.4. *A* and *B* Canonical Analysis for APL1P

Most importantly, the canonical transformations of (4.51-54) can provide an algebraic description compatible with the graphical depiction of Figure 4.4. Again, adapting Box and Draper's (1987) general description of canonical analysis to the recourse problem shows

1. *Eigenvalues*. The eigenvalues λ^k measure the degree of slope in the transformed coordinate system $\mathbf{X}$, while the signs indicate the type of contour surface (i.e., concave, convex, rising, saddle, etc.). *Since the literature has shown $Z(\mathbf{x})$ is either a convex (minimization) or concave (maximization) function of $\mathbf{x}$, λ must be either strictly positive (minimization) or strictly negative (maximization).* Thus the eigenvalues can not tell only the relative *sensitivity* of the fitted response $V_A(\mathbf{X})$ to movement along the rotated axes, but can also confirm the *validity* of the fit through their signs in the context of the recourse problem. In the case of APL1P, both eigenvalues are significant, although curvature along the rotated X^2 (996.5) axis is four times as steep as that of X^1 (229.8).

2. *Eigenvectors*. The normalized eigenvectors of $\mathbf{M}$ describe the component contribution of the original $\mathbf{x}_K$ axes to the rotated axes $\mathbf{X}$ through the relationship $\mathbf{X} = \mathbf{M}^T\mathbf{x}_K$. In practical terms, this means the elements of the eigenvectors provide the basis for estimating the factor dependence — or tradeoffs — of the decision variables $\mathbf{x}_K$. For instance, in the case of APL1P a roughly equal presence of both decision variables x^1 and x^2 — (.7396,-.6730) and (.6730,.7396), respectively — indicate an approximate 45^0 rotation of the fitted quadratic surface to the original decision variables $\mathbf{x}$.

3. *Slope*. The variable θ^k measures the slope of the fitted response in the direction of the rotated axis X^k. For APL1P X^1 has a far higher linear

coefficient than X^2. In conjunction with the stronger curvature of X^2, this indicates that, while $V_A(\mathbf{X})$ is less sensitive to changes in X^1 than in X^2, there does not exist a *stationary ridge* through the entire design region.

4. *Stationarity*. The distance D of the design origin to the estimated stationary (minimum) point is

$$D = \left\{ \sum_{k=1}^{K} (X_s^k)^2 \right\}^{1/2} \tag{4.55}$$

and provides a measure of how close the optimum point of the fitted response $V_A(\mathbf{X})$ is to the center of the experimental design $\mathbf{x}^*$ or $\mathbf{x}'$ (Box and Draper 1987). In the case of APL1P $D = [(-.4658)^2 + (-.0146)^2]^{1/2} = .466$, as seen in Figure 4.4.

For closely located stationary points and design centers, a second canonical form — the *B Canonical Analysis* — can further simplify the polynomial approximation (4.53) by centering the rotated axes $\mathbf{X}$ around the stationary point $\mathbf{X}_s$.

4.4.3 *B* Canonical Form

Box and Draper (1987) suggest employing the *B* canonical analysis whenever the experimental design and stationary point of the fitted surface are approximately the same (i.e., the fitted surface closely approximates the true response), and suggest $D \leq 1$ as a benchmark. This transformation occurs using the relationship

$$-2\Lambda\mathbf{X}_s = \theta \tag{4.56}$$

which gives the fitted response

$$\hat{Z}(\mathbf{X}_s) = a_0 + 1/2\mathbf{X}_s'\theta. \tag{4.57}$$

Defining $\tilde{X}^k$ as

$$\tilde{X}^k = X^k - X_s^k \tag{4.58}$$

and its corresponding vector $\tilde{\mathbf{X}} = [\tilde{X}^1, \ldots, \tilde{X}^K]^T$ produces the *B* canonical form of (4.48)

$$V_B(\tilde{\mathbf{X}}) = \hat{Z}(\mathbf{X}_s) + \tilde{\mathbf{X}}'\Lambda\tilde{\mathbf{X}} \tag{4.59}$$

and of (4.49)

$$V_B(\tilde{\mathbf{X}}) = \hat{Z}(\mathbf{X}_s) + \sum_{k=1}^{K} \lambda^k(\tilde{X}^k)^2 \tag{4.60}$$

where $V_B(\mathbf{X} - \mathbf{X}_s) = V_A(\mathbf{X})$ (Box and Draper 1987). For the example problem APL1P, the *B* canonical form becomes

$$V_B(\tilde{\mathbf{X}}) = 24543 + 229.8(\tilde{X}^1)^2 + 996.5(\tilde{X}^2)^2 \tag{4.61}$$

as indicated by the dashed lines in Figure 4.4.

4.4.4 Ridge Analysis

The final aspect of RSM employed by this dissertation provides the best insight into the recourse problem — the direction of minimum and maximum

sensitivity of the fitted response $V(\mathbf{x}_K)$ to changes in $\mathbf{x}_K$. Again paraphrasing Box and Draper (1987), the idea starts with the assumption that for the fitted response there exists p eigenvalues whose small values imply a p-dimensional ridge. Indexing the eigenvalues from largest to smallest gives $\lambda = [\lambda^1, \ldots, \lambda^{K-p}, \lambda^{K-p+1}, \ldots, \lambda^K]^T$ and their respective rotated axes $\mathbf{X} = [X^1, \ldots, X^{K-p}, X^{K-p+1}, \ldots, X^K]^T$. The distance DR from the design centerpoint to the ridge is then

$$DR = \left\{ \sum_{k=1}^{K-p} (X_s^k)^2 \right\}^{1/2} \tag{4.62}$$

and the coordinates of the rotated system nearest the ridge are $\mathbf{X}_s = [X_s^1, \ldots, X_s^{K-p}, 0, \ldots, 0]^T$. Proceeding from $\mathbf{X}_s$ and moving exactly one unit in either direction of X^K provides three sample responses

$$(1)\ \hat{Z}(\mathbf{X}_s), \quad (2)\ \hat{Z}(\mathbf{X}_s) + \theta^K + \lambda^K, \quad \text{and } (3)\ \hat{Z}(\mathbf{X}_s) - \theta^K + \lambda^K, \tag{4.63}$$

whose average and sample variance are

$$\hat{Z}(\mathbf{X}_s) + 2/3 \cdot \lambda^K \tag{4.64}$$

and

$$(\theta^K)^2 + 1/3 \cdot (\lambda^K)^2, \tag{4.65}$$

respectively. Since the standard deviation of a normal distribution can be estimated with three samples using the relationship

$$\frac{\textit{sampling range}}{3^{1/2}}, \tag{4.66}$$

the approximate sample range r^K for $\mathbf{X}_s$ in the direction of the rotated axis X^K within the CC portion of the experimental design is

$$r^K = [3(\theta^K)^2 + (\lambda^K)^2]^{1/2} \tag{4.67}$$

(Box and Draper 1987).

Applying (4.67) to APL1P produces

$$r^1 = [3\cdot(214.09)^2 + (229.8)^2]^{1/2} = 436.25 \tag{4.68}$$

for X^1, and

$$r^2 = [3\cdot(29.19)^2 + (996.5)^2]^{1/2} = 997.78 \tag{4.69}$$

for X^2. Results (4.68-69) algebraically confirm what Figure 4.4 shows as the minimum sensitivity to be along the rotated X^1 axis, and to expect a $Z(\mathbf{x})$ of no more than 24643 + 436 = 25079 within plus or minus one unit distance from $\mathbf{X}_s$. Equally important, these results show that movement along the X^2 axis provides the *worst* deviation from the optimal solution, and thus should be avoided. Since the original design center point essentially lies on the X^1 axis, the analysis can effectively conclude that near-optimal solutions lie on a ridge defined by the relation

$$\mathbf{x} = (1800,1570) + \rho\cdot(.7396,-.6730);\ \rho \in [-1082,+1082]. \tag{4.70}$$

This relationship can be presented to the decision-maker — in even simpler terms — to provide the basic insight into the problem solution.

APL1P Response Analysis Summary. *A near-optimal solution requires (1) a total combined investment of 3370 for x^1 and x^2; (2) a one-for-one tradeoff between x^1 and x^2 starting at (1800,1570); and (3) neither $x^i <$ 1000.* ***Generally speaking, the total amount of installed capacity is more important than how it is split between the generators.***

4.5 DISTRIBUTION ANALYSIS

4.5.1 Introduction

McKay (1992) cites two primary questions asked about the uncertainty of the output of any simulation model: *'What Influences It?'* and *'How Large?'*. The previous section shows how RSM answers the first question regarding which decision variables affect the response and by how much. Typically, such analysis also presents the decision-maker with a range of multiple optimal or near-multiple optimal solutions due to the 'flatness' of the region of optimality, thus allowing the use of subjective criteria and individual judgment not captured in the original model. *However, for stochastic linear programming problems, McKay's second question remains unanswered.* Until now, both the literature and this dissertation focus on the expected value of the recourse function (3.1b), assuming it to be the primary decision criteria (in conjunction with known first-stage costs). Instead, this dissertation contends most decision-makers want a ***range*** of possible outcomes based on their decision — not just an average. This thesis also contends that such knowledge will narrow the choice of solution in near-optimal situations by comparing of the range of possible outcomes for each **x** and selecting the one with the smallest variance, best worst-case scenario, best best-case scenario, or

some other criteria. *In effect, it introduces the underlying distribution as a decision criteria for two-stage stochastic linear programming problems with recourse.*

Incorporating the variability of $h(\mathbf{x},\omega,\mathbf{T})$ in the modeling process represents this dissertation's other major contribution to this area of research. This section inaugurates this topic by reviewing its distributional features, then outlining the measures of uncertainty used to characterize it. It closes the chapter by applying its techniques on the example problem APL1P and proposing a final solution recommendation.

4.5.2 Distributional Characteristics

Deciding which measure of uncertainty to employ depends upon the attributes (or assumptions) regarding the underlying distribution. Redesignating the k subscript to denote a distinct solution $\mathbf{x}_k$, recall the definition of z_{ik} as the optimal value of the recourse function (3.1b) plus $\mathbf{cx}_k$ for the i^{th} realization of ω and $\mathbf{T}$; and, the random variable z_k distributed as $\mathbf{cx}_k + h(\mathbf{x}_k,\omega,\mathbf{T})$, where $E[z_k] = Z(\mathbf{x}_k)$. Since the distribution of z_k is a function of $\mathbf{x}$, its form and parameters will vary throughout the feasible region X, and must be estimated for larger problems. In the case of smaller problems where the population distribution can be determined for $\mathbf{x}_k$, then statistical estimation is not required. Most importantly, it is very unlikely that z_k will follow a unimodal or symmetric distribution.

Empirical evidence from APL1P suggest this asymmetry occurs even for the simplest problems. Using the same $\mathbf{x}_k$ points from Table 4.2, Table 4.5 lists

TABLE 4.5
DISTRIBUTION PARAMETERS FOR VRT SAMPLES IN TABLE 4.2 FOR APL1P

$\mathbf{x}_k$	$\bar{z}_k = Z(\mathbf{x}_k)$	Median z_k	S. D. of z_k	Min z_k	Max z_k
(1800,1800)	24689.1	26160.0	4808.2	18270.0	45990.0
(900,900)	26425.4	27705.0	3553.5	17550.0	40995.0
(900,2700)	25131.3	25265.0	5207.5	18720.0	45495.0
(2700,900)	25299.3	25072.0	5282.0	19170.0	46485.0
(2700,2700)	27499.3	27210.0	4070.5	23670.0	50985.0

the population mean, median, and standard deviation; and, the lowest and highest possible values for z_k. The differences in the mean and median of z_k suggest varying degrees of symmetry of the underlying distribution, while the differences in standard deviations indicate differences in dispersion as well. This lack of symmetry does not, of course, mean that standard statistical analysis techniques employing the central limit theorem, such as confidence interval estimation regarding $\hat{Z}_s(\mathbf{x})$, do not apply. Furthermore, the literature does not suggest that z_k itself is normally distributed, or should even follow a symmetric distribution. It does mean, however, that any *assumptions made by the decision-maker* regarding the distributional form of z_k can be misleading.

Figures 4.4 and 4.5 graphically show the advantages of more accurate distributional analysis, and illustrate the skewed nature of z_k, by comparing population histograms for the points (1800,1800) and (2700,2700), respectively, to normal distributions *with the same mean, variance, and area* (based on a presentation idea by Bradley 1968). Both samples share — in varying degrees —

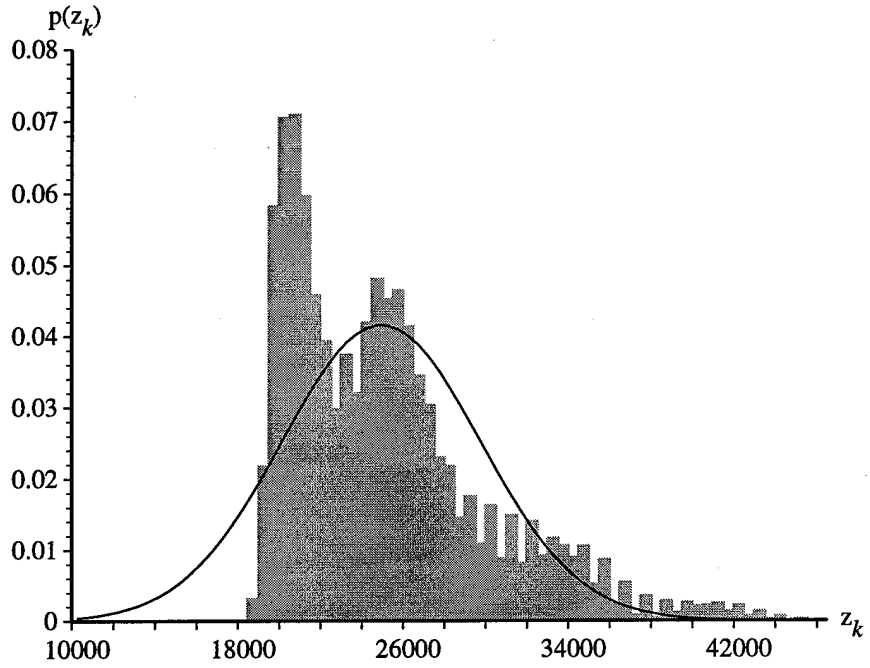


Figure 4.5. Comparison of Population Distribution of (1800,1800) in APL1P to Normal Distribution with Same Mean, Variance, and Area

a heavily skewed dispersal favoring the area just below the average, an attenuated trimodal shape, and a significant presence of high-cost solutions far above what occurs in the upper tail of a normal distribution. Clearly, any assessment using normal or symmetric assumptions on the part of the decision-maker would not be appropriate in either case — such presuppositions would overstate the frequency of low-cost solutions while understating the high ones. Additionally, for non-symmetric distributions in general, the *variance* of z_k does not adequately characterize its *coverage* (McKay 1992).

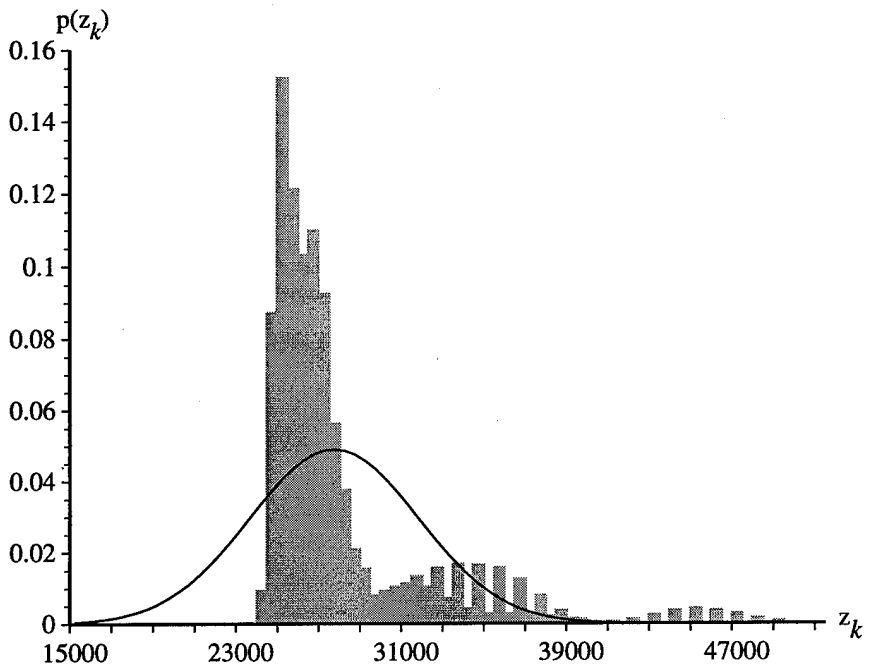


Figure 4.6. Comparison of Population Distribution of (2700,2700) in APL1P to Normal Distribution with Same Mean, Variance, and Area

Although difficult to predict in its exact form, intuitively it is not surprising that the distribution of z_k possesses such characteristics. As an optimization function, it would shift bases to minimize changing resource demands in ω_i - $\mathbf{T}_i\mathbf{x}_k$, thus mitigating somewhat their effects on the cost. Furthermore, marginal increases in ω_i - $\mathbf{T}_i\mathbf{x}_k$ over ω_j - $\mathbf{T}_j\mathbf{x}_k$ ($j \neq i$) may occur in slack resources that would have no effect in increasing the solution value z_{ik} from z_{jk}; thus, the 'bunching' effect observed in Figures 4.5 and 4.6. Consequently, higher values of z_k should occur disproportionately fewer times than the distribution of the elements in ω and $\mathbf{T}$ would suggest.

Based on empirical evidence as seen in Figures 4.5 and 4.6, and in additional preliminary research, this dissertation contends that non-parametric, or distribution-free, statistical analysis presents an excellent option for analyzing the range of coverage of z_k for reasons of simplicity, computational efficiency, and minimal assumptions of the underlying distribution (see Bradley 1968). Consequently, following Wilson's (1995) suggestion, this research investigates the distribution of $h(\mathbf{x},\omega,\mathbf{T})$ by employing *tolerance limits* as the basic measure of interest for z_k. As a non-parametric statistic, it possesses characteristics that make its application appealing in the context of the stochastic recourse problem. More importantly, it provides the necessary information about the dispersal of z_k; and, in conjunction with the response surface approximation of $Z(\mathbf{x})$, gives the decision-maker a truly useful description of the stochastic behavior of (3.1).

4.5.3 Tolerance Limits

Quoting Conover

> ... confidence intervals ... provide interval estimates for unknown population parameters, such as the unknown probability p or the unknown quantile x_p, and a certain probability $1 - \alpha$ (confidence coefficient) that the unknown parameter is within the interval. Tolerance limits differ from confidence intervals in that tolerance limits provide an interval within which at least a proportion q of the *population* [emphasis added] lies, with probability $1 - \alpha$ or more that the stated interval does indeed "contain" the proportion q of the population (Conover 1980, 117).

Thus, by estimating the population range tolerance limits provide a measure of 'coverage' for asymmetrical distributions that convey a useful characterization of

the distribution of z_k to the decision-maker. Adapting Conover's (1980) notation and terminology to the recourse context gives the following definitions.

> **Definition.** Let z_i represent $\mathbf{cx} + h(\mathbf{x},\omega_i,\mathbf{T}_i)$ for the i^{th} realization of a random sample of ω and $\mathbf{T}$, $i = 1, \ldots, I$ (again, dropping the k subscript for convenience).

> **Definition.** Let the parameters r and m, where $r < m$, index the ordered sample $z_1 \leq \ldots \leq z_r \leq \ldots \leq z_m \leq z_I$, and $1 \leq r < m \leq I$. Furthermore, let $z_0 = -\infty$ and $z_{I+1} = +\infty$.

Conover describes the tolerance interval approach as determining the sample size I such that for a probability of at least $1 - \alpha$ no less than a q portion of the population lies between z_r and z_{I+m-1}, where q, r, m, and α are predetermined. Note that this formulation allows either one sided tolerance limits (r or m equals zero) or a two-sided tolerance intervals (r and m not equal to zero), and can be approximated with the relationship

$$I \approx \left\{\frac{\chi_{1-\alpha}}{4} \cdot \frac{1+q}{1-q}\right\} + \frac{r+m-1}{2} \tag{4.71}$$

where $\chi_{1-\alpha}$ is the $(1 - q)$ quantile of a chi-squared distribution based on $2\cdot(r + m)$ degrees of freedom (Conover 1980).

Examining (4.71) shows that the required sample size increases the most for higher proportions of the coverage percentage q, and less so for increases in the confidence level α and indices r and m. Striking a balance between these

factors and the computational requirements of sampling $h(\mathbf{x},\omega,\mathbf{T})$, this dissertation will set $r = 1$ and $m = 1$, $q = .05$, and $\alpha = .01$, giving a sample size $I = 130$. Note that this sample size does not depend upon the distribution's form or size, and thus can be applied for any stochastic recourse problem.

4.5.4 Example Distribution Analysis

Demonstrating the tolerance limits method on APL1P, Table 4.6 shows the results for five equidistant $\mathbf{x}_k$ taken along the minima ridge defined by (4.70). First, note that the probability coverage meets or exceeds 95% for all points except (1000,2298), which falls slightly below the targeted proportion q. Second, in general the tolerance limits provide a very good approximation of the range for both individual estimates of $\mathbf{x}_k$ and comparisons across the ridge. For instance, the tolerance limits' upward trend of the estimated maximum z_k with respect to increases in ρ matches the actual upward trend of the population maximum. Third, both the tolerance limits and $Z(\mathbf{x})$ tend to favor the minima ridge from the optimal point to the stationary point X_S of the response surface approximation. Fourth, there does exists a difference between the population extremes and those found by tolerance limit sampling. However, the purpose of tolerance limits is to estimate the *range* we can expect z_k to fall into for a q portion of the time — not to provide *point* estimators. Finally, Figures 4.7 and 4.8 graphically compare the tolerance limits data in Table 4.6 to the distribution of z_k for the two best sampled points with respect to $Z(\mathbf{x})$ (i.e., (1400,1934) and (1800,1570)).

TABLE 4.6
TOLERANCE LIMITS FOR APL1P MINIMA RIDGE (RANDOM SEED = 674303674)

				Population		Tol. Limit		
ρ	x_k	Z(x)	Mdn. z_k	Min z_k	Max z_k	Min z_k	Max z_k	% Cvg.
-1082	{1000,2298}	24828	25493	17915	44795	19193	38727	.9493
-541	{1400,1934}	24689	26285	17805	45105	18405	40364	.9864
0	{1800,1570}	24642	26075	17695	45415	18695	41002	.9610
541	{2200,1206}	24801	25415	18185	45725	18535	41725	.9861
1082	{2600,842}	25230	25093	18675	46035	18925	43035	.9921

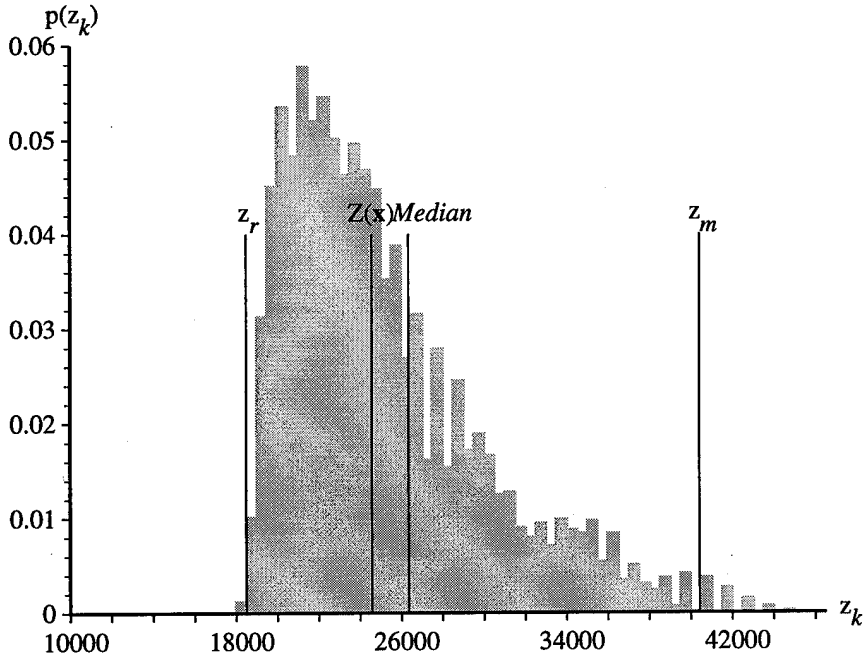


Figure 4.7. Comparison of Tolerance Limits to Population Distribution, Mean, and Median for (1400,1934) in APL1P

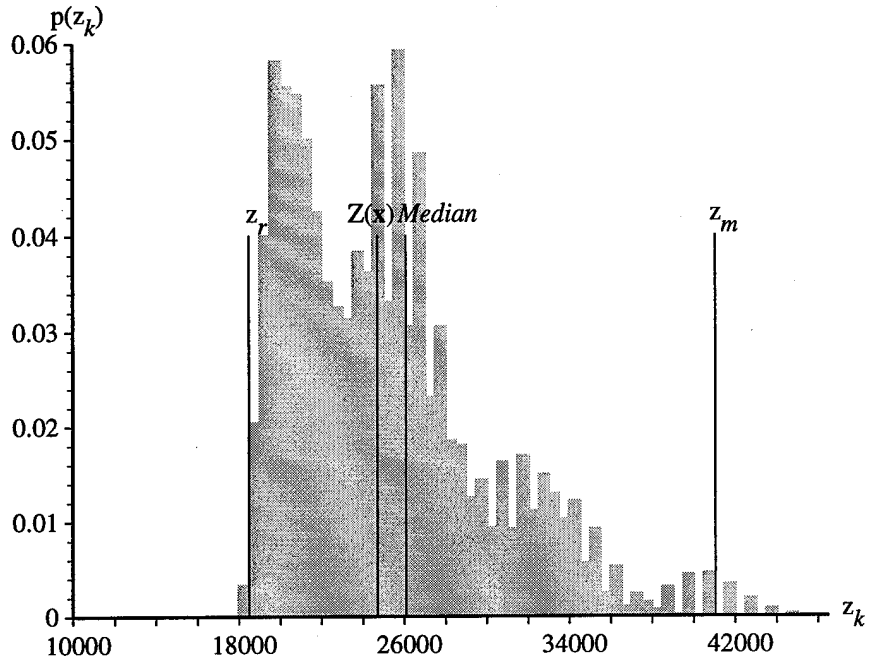


Figure 4.8. Comparison of Tolerance Limits to Population Distribution, Mean, and Median for (1800,1570) in APL1P

These sample points indicate that $\mathbf{x}_k = (1400,1934)$ offers a lower range of possible z_k values without sacrificing much of its expected value. Based on these results, the distributional analysis suggests restricting the final choice by modifying (4.70) to be

$$\mathbf{x} = (1400,1934) + \rho\cdot(.7396,-.6730),\ \rho \in [0,541], \tag{4.72}$$

and presenting to the decision-maker the following guidance.

APL1P Distributional Analysis Summary. *A near-optimal solution requires (1) a total combined investment of 3350 ± 20 for* $\mathbf{x}$*; and, (2) a*

one-for-one exchange between x^1 and x^2 starting at (1400,1934) and ending at (1800,1570). Furthermore, as the solution moves away from (1400,1934) expect a tradeoff between slightly lowering the expected cost and marginally increasing the expense of the worst-case scenario.

Finally, it should be noted that using distribution-free statistics like tolerance limits does not diminish either the utility of basing the response surface on $Z(\mathbf{x})$. First — and most significantly — both the search techniques and the response surface analysis fundamentally depend on the convexity of $Z(\mathbf{x})$. Thus, changing the search and approximation criteria away from $Z(\mathbf{x})$ poses significant challenges that are beyond the scope of this research. Second, intuitively it appears likely that the range of z_k will be a function of $\mathbf{x}$; for example, the lowest *maximum* value of z_k for the *entire* set $\boldsymbol{X}$ will probably occur somewhere in or near the region of optimal or near-optimal solutions. Therefore, using $\mathbf{x}^*$ or $\mathbf{x}'$ as a starting point for the response surface and distributional analysis can still be justified. Finally, this research wants to *supplement* the use of $Z(\mathbf{x})$, not *replace* it; consequently, distributional analysis presents a natural and obvious method for providing the decision-maker with additional insight into stochastic linear programming problems with recourse.

Chapter 5

Problem Set Analysis

5.1 INTRODUCTION

This chapter demonstrates the response surface analysis approach outlined in Chapters 3 and 4 on the set of problems listed in Table 5.1. The subsequent sections each cover a single problem by providing a brief formulation and description followed by an analysis similar in structure and content to the one presented for APL1P (deviations from the prescribed algorithms are noted as well). These problems have also been independently solved (Morton 1994c), thus providing a benchmark for confirming the optimal solutions. Table 5.1 also shows which optimization and analysis techniques from Chapters 3 and 4 are used for each problem.

As in the case of APL1P, the computational environment consists of an IBM RS/6000 Model 320 running under AIX 3.2 and FORTRAN 90 for the OBS and RSA programs, while SAS Version 6.08 running on OpenVAX/VMS provides the response surface analysis. As before, all timing results are based on AIX's estimate of CPU code execution, and do not include system overhead or I/O requirements. The code also requires linking to IBM's OSL and IMSL libraries. Finally, due to the size of some problems this chapter depicts only the relevant portions of the formulation.

TABLE 5.1
APPLICATION SUMMARY OF OPTIMIZATION AND STATISTICAL ANALYSIS TECHNIQUES

	Problem Set Description				
	APL1P	PGP2	CEP1	4TERM	20TERM
# of **x** Variables	2	4	4	15	63
Rows/Columns in **A**	2 / 2	2 / 4	9 / 8	3 / 15	3 / 63
Rows/Columns in **W**	5 / 9	7 / 16	7 / 15	28 / 146	124 / 764
# of Random Variables in ω/**T**	5	3	3	8	40
# of Scenarios	1280	576	216	256	$1.095 \cdot 10^{12}$
	Optimization				
	APL1P	PGP2	CEP1	4TERM	20TERM
Geometric Simplex	Yes	Yes	Yes	Yes	—
Projected Gradient	Yes	Yes	Yes	Yes	Yes
PARTAN	Yes	Yes	Yes	Yes	—
OSL	Yes	Yes	Yes	Yes	Yes
OBS-COMPLETE / ODV	Yes	Yes	Yes	—	—
OBS-RESET	—	—	—	Yes	—
	Statistical Analysis				
	APL1P	PGP2	CEP1	4TERM	20TERM
Control Variates	Yes	Yes	Yes	Yes	Yes‡
Latin Hypercube	Yes	Yes	Yes	Yes	Yes
Population†	Yes	Yes	Yes	Yes	—
Full Exper. Design	Yes	Yes	—	—	—
Fractional Exper. Design	—	—	—	Yes	Yes
Prelim. Screening Design	—	—	—	—	Yes
Resp. Surf. (Minima Ridge)	Yes	Yes	—	Yes	—
Resp. Surf. (Maxima Ridge)	Yes	Yes	—	Yes	Yes
Tolerance Limits	Yes	Yes	Yes	Yes	Yes

† - 'Population' implies all scenarios evaluated for validation of VRTs and Tol. Limits. Additionally, true expected values used in experimental design when available.
‡ - Used only for comparison with LHs; NOT used for reported response surface estimates.

5.2 PGP2

5.2.1 PGP2 Problem Description

PGP2 represents a power generation expansion problem developed by Louveaux and Smeers (1988) and modified for use by the University of Michigan FTP site (Holmes 1995). Figure 5.1 on the following page gives its formulation, whose recourse configuration contains the transportation problem structure found

MIN	$10x^1$	$7x^2$	$16x^3$	$6x^4$	$40y^1$	$24y^2$	$4y^3$	$45y^4$	$27y^5$	$4.5y^6$	$32y^7$	$19.2y^8$	$3.2y^9$	$55y^{10}$	$33y^{11}$	$5.5y^{12}$	$1kz^1$	$1kz^2$	$1kz^3$	$1kz^4$	
S.T.	x^1	x^2	x^3	x^4																	≥ 15
	$10x^1$	$7x^2$	$16x^3$	$6x^4$																	≤ 220
	$-x^1$				y^1	y^2	y^3										$-z^1$				≤ 0
		$-x^2$						y^4	y^5	y^6								$-z^2$			≤ 0
			$-x^3$								y^7	y^8	y^9						$-z^3$		≤ 0
				$-x^4$										y^{10}	y^{11}	y^{12}				$-z^4$	≤ 0
					y^1			y^4			y^7			y^{10}							$\geq \omega^1$
						y^2			y^5			y^8			y^{11}						$\geq \omega^2$
							y^3			y^6			y^9			y^{12}					$\geq \omega^3$

Figure 5.1. Formulation of PGP2 Problem
('+' signs omitted for clarity. $1k = 1000$.)

in many capacity expansion problems. One difference from APL1P lies with the first-stage constraints; in PGP2's case, $\mathbf{Ax} = \mathbf{b}$ includes both a minimum supply capacity and a first-stage capital improvement budget constraint. PGP2 also limits its source of variability to the right-side vector ω of the recourse problem, whose discrete representation allows for 576 possible demand scenarios. Finally, the recourse error vectors z^j (with objective coefficient values of 1,000) ensure feasibility for all possible values of $\mathbf{x}$ and ω.

5.2.2 PGP2 Optimization Results

The OBS algorithm found 32 optimal bases and their respective dual vectors for PGP2 (Table 5.2), with a frequency of occurrence listed in Table 5.3. The OBS-COMPLETE algorithm subsequently found one additional basis during the line search portion of the PARTAN search. The frequency distribution of the optimal basis set for PGP2 shows less concentration of optimality than seen with APL1P; e.g., 92% of optimality occurs under bases #1 through #8 for APL1P versus 74% for PGP2. This greater dispersion, combined with the larger number of dual vectors, results in the OBS-COMPLETE and ODV algorithms turning in comparable computation times for all three search techniques (Table 5.4). Both options turn in performance times an order of magnitude better than OSL alone.

Due to the relatively small number of scenarios in PGP2, all search algorithms calculate the exact value of $Z(\mathbf{x})$ for this problem. In turn, this lack of experimental error most likely causes the GEOMETRIC SIMPLEX ALGORITHM to

TABLE 5.2
OBS-COMPLETE RESULTS FOR PGP2

Sample Size (ω - **Tx**)	Random # Seed	All Bases / First Optimal*	# Opt. Bases/ Dual Vectors	CPU Time (secs)
1,000	3203801	All	21	15.34
5,000	7733099	All	25	33.24
50,000	11351	First	31	161.95
250,000	603939541	First	32†	229.48

* - 'First Optimal' option skips any remaining bases after finding first feasible, whereas 'All Bases' checks every basis in P for each sample (ω - **Tx**)
† - Subsequent search routines found 1 additional optimal basis.

TABLE 5.3
FREQUENCY OF BASIS OPTIMALITY FOR PGP2
(BASED ON 4th RUN FROM TABLE 5.2)

Basis ID #	1	2	3	4	5	6	7	8	9	10	11	12	13-32
Freq. of Optimality	.19	.19	.09	.09	.06	.05	.04	.03	.03	.03	.02	.02	.14
Cumulative Freq.*	.19	.38	.47	.56	.62	.67	.71	.74	.78	.81	.83	.86	1.0

* - May not add due to roundoff error.

TABLE 5.4
COMPUTATION TIMES OF OSL/OBS/ODV OPTIONS FOR GEOMETRIC SIMPLEX, PROJECTED GRADIENT, AND PARTAN ALGORITHMS FOR PGP2 (IN SECONDS)

Algorithm	OSL	OBS*	ODV
GEOMETRIC SIMPLEX (120 Iterations)	861.67	96.3	87.62
PROJECTED GRADIENT (22 Iterations)	1028.41	66.24	65.99
PARTAN (2 Iterations)	212.22	14.75	13.54

* - PARTAN search found one additional optimal basis.

contract on its best vertex. As shown in Table 5.5, the algorithm initially progresses beyond its starting set of vertex points, but by iteration 40 settles into slowly contracting about its best vertex. At k = 86, it further shrinks by

contracting, then flipping about the best point and enlarging when further improving moves are not possible. Another contraction follows the enlargement move when *it* does not provide any better vertices; and, when the final contraction offers no help either, the simplex re-initializes itself and proceeds with iteration 86. Having kept the best vertex from the previous simplex set, the procedure settles back into its slow contraction pattern by $k = 100$ with little improvement in the objective function value. This premature contraction most likely occurs since

TABLE 5.5
SELECTED GEOMETRIC SIMPLEX MOVES FOR PGP2
(RANDOM SEED = 296279)

k	x^1, x^2, x^3, x^4	Z(x)	Simplex Move	Replaces
0	4.00, 0.00, 5.00, 6.00	504.40	Initial Vertex ($\mathbf{x}_{ev}$)	—
1	6.99, .77, 5.18, 6.45	466.40	Expansion $\mathbf{x}_e$	5
3	5.47, 0.84, 5.94, 4.38	455.02	Contraction $\mathbf{x}_c$	5
10	5.10, 1.29, 6.12, 4.96	453.84	Contraction $\mathbf{x}_c$	5
20	2.76, 1.99, 5.69, 7.35	450.39	Contraction $\mathbf{x}_c$	5
40	3.02, 1.98, 5.88, 6.77	449.94	Contraction $\mathbf{x}_c$	5
85	3.12, 1.93, 5.88, 6.69	449.88	Contraction $\mathbf{x}_c$	5
85-86	—	—	Shrinkage	—
85-86	—	—	Enlargement	—
85-86	—	—	Shrinkage	—
85-86	—	—	New Simplex	—
86	7.79, 1.95, 3.13, 8.24	475.74	Expansion $\mathbf{x}_e$	5
90	4.64, 2.44, 5.44, 4.07	452.16	Expansion $\mathbf{x}_e$	5
100	3.31, 1.70, 5.86, 6.48	449.68	Contraction $\mathbf{x}_c$	5
110	1.54, 2.19, 6.64, 7.10	449.42	Expansion $\mathbf{x}_e$	5
120	2.70, 1.81, 6.12, 6.82	449.32	Contraction $\mathbf{x}_c$	5

every interior point of the simplex offers a slight improvement over the next-worst vertex value; however, the algorithm does offer considerable improvement over the expected value approximation $Z(\mathbf{x}_{ev})$.

The PROJECTED GRADIENT ALGORITHM (Table 5.6) performs very well in the case of PGP2. Since it does not encounter either a binding constraint set or a multiple optimal solution region, the algorithm suffers from not finding a zero gradient due to the non-differential property. Consequently, terminating the search requires an arbitrary stopping point γ that in this case was determined by feedback from the first set of iterations. After initially finding a good solution at step 9 (based on $d^j_k < -1.0$, $j = 1,\ldots, 4$), the search 'overshoots' the optimal solution in step 10 as evidenced by $d^3_k = -1.71$, even though $Z(\mathbf{x}_{10}) < Z(\mathbf{x}_9)$. Therefore, by setting $\gamma^j = -.99$, $j = 1,\ldots, 4$, the search continues the descent until reaching step 22 where $\mathbf{d}_k$ again is less than -.99 for each of its components. Finally, the fact that solutions better than step 11's do not occur in steps 13 through 20 tend to confirm $\mathbf{x}_{22}$ as a near-optimal solution.

Based on the results of the PROJECTED GRADIENT ALGORITHM, the PARTAN search (Table 5.7) stops after reaching a similar solution with respect to $Z(\mathbf{x})$ and $\mathbf{d}_k$, using this prior knowledge allows the search to conclude after two iterations. One notable difference between this result and the proposed algorithm in Chapter 3 concerns the differences between the estimated optimal scalar multiple and the one actually used. In PGP2, the quadratic fit of the line spanning the entire feasible region (based on $\mathbf{x}_{k-1}$ and $\mathbf{p}_k$) rarely exceeds a R^2 of .80. After several preliminary trials, a subjective interpolation of the data points proved to

TABLE 5.6
SELECTED PROJECTED GRADIENT ITERATIONS FOR PGP2
(SCENARIOS = 576, $Q = 8$)

k	x^1, x^2, x^3, x^4	$\mathbf{d}_k$	Est. q	Act. q	R^2	Z(x)
1	5.00, 5.00, 5.00, 5.00	8.70, 6.95, 8.97, 5.99	0.42	.42	.900	466.62
2	3.92, 4.37, 4.88, 4.74	6.65, 5.77, 5.65, 5.13	0.24	.24	.995	451.61
3	3.56, 4.08, 5.03, 4.54	0.58, -0.30, 0.85, 0.86	0.03	.03	.998	449.13
4	3.44, 4.16, 4.91, 4.76	0.58, -0.30, -0.43, -0.86	0.28	.28	.971	447.79
8	2.35, 4.32, 5.07, 5.53	-0.71, -0.95, -0.43, -0.85	-2.11	.01	.999	447.78
9	2.37, 4.40, 5.02, 5.59	-0.71, -0.95, -0.43, -0.85	-2.48	.01	.999	447.67
10	2.38, 4.47, 4.97, 5.65	-0.71, -0.95, -1.71, -0.84	-0.98	.005	.999	447.60
11	2.37, 4.48, 4.99, 5.65	-0.70, -0.94, -1.71, -0.84	-1.04	.005	.999	447.56
21	2.26, 4.57, 5.00, 5.66	-0.71, -0.95, -1.71, -0.84	-0.93	.002	.999	447.56
22	2.25, 4.57, 5.00, 5.65	-0.71, -0.95, -0.43, -0.84	-2.95	—	.999	447.55

TABLE 5.7
PARTAN ITERATIONS FOR PGP2 (SCENARIOS = 576, $Q = 8$)

k	x^1, x^2, x^3, x^4	Est. q^*	Act. q	R^2	Z(x)
$\mathbf{x}_0$	5.00, 5.00, 5.00, 5.00	.53	.71	.743	466.62
$\mathbf{x}_1$	3.57, 4.17, 4.84, 4.65	.71	.71	.985	449.06
$\mathbf{p}_1$	2.60, 4.29, 5.10, 5.47	.52	.70	.767	447.90
$\mathbf{x}_2$	2.43, 4.24, 5.11, 5.50	.48	—	.791	447.84

give better results than following the regression's estimate. Consequently, the results reflect this deviation.

5.2.3 PGP2 Response Surface Analysis

A preliminary full-factorial CCD experimental design using $\mathbf{x}_0$ =(2.25, 4.55, 5.00, 5.50) as the centerpoint and each factor's half-range consisting of ±0.5

gives a semi-definite fit (i.e., a saddle point) based on the eigenvalues and eigenvectors reported in Table 5.8 (albeit the three negative eigenvalues are only marginally curved downward). Since this obviously represents an incorrect fit based on the known convexity of the response, the final design departs from the standard CCD guidance to induce a positive-definite outcome. Based on the large contributions of x^2, x^3, and x^4 to those rotated axes with downward curvature, the subsequent design drops the axial representation of x^1, and extends those of x^2, x^3, and x^4 enough to leverage the regression into the correct fit. (The alternative approach of increasing the size of the CCD would impose asymmetrical axial points to remain feasible.) Additionally, the minima ridge estimates from the preliminary design provided a slightly better centerpoint location for the final design for PGP2 ($\mathbf{x}_0$ = (2.271, 4.605, 5.045, 5.567)). Although these modifications would normally destroy the uniform-precision or rotatibility of the design, in this case error resulting from response variability does not occur due to using the exact values of $Z(\mathbf{x})$. Consequently, the bias inherent in a polynomial approximation constitutes the sole source of any lack of fit, and thus helps mitigate the effects of altering the design.

TABLE 5.8
A CANONICAL ANALYSIS OF PGP2 (PRELIMINARY CCD)

	Eigenvectors			
Eigenvalues	x^1	x^2	x^3	x^4
25.4760	0.5064	0.4947	0.5190	0.4791
-0.1255	-0.0222	0.3542	-0.7849	0.5080
-0.5921	-0.7749	-0.0720	0.3337	0.5319
-0.7009	0.3776	-0.7904	-0.0573	0.4790

Table 5.9 gives the final experimental design, Table 5.10 provides the regression parameter estimates, and Table 5.11 supplies the eigenvalue, eigenvector, and ridge results for PGP2. Although not readily apparent, the very small differences between the Z(**x**)s of the fractional portion of the design with the centerpoint's Z(**x**) give some indication of how a semi-definite fit can occur without substantially higher axial response values. Tables 5.10 and 5.11 confirm a good fit with a high R^2 and positive eigenvalues, respectively. Most importantly, Table 5.11 provides the minima and maxima ridge analysis associated with the reported eigenvalues and eigenvectors.

The *A* canonical analysis in Table 5.11 shows decision variables x^2 and x^3 as roughly equal components in the rotated axis containing the highest amount of curvature; by contrast, x^1 and x^4 dominate those rotated axes with the least increase in the response Z(**x**) per unit change from the centerpoint. Consequently, the minima ridge in the original coordinate system occurs along a vector where increases in x^2 and x^3 are kept to a minimum by lowering x^1 and x^4. Reversing the criteria for the maxima ridge produces a vector with a rapid rise in the two major constituent variables of the first eigenvector (x^2 and x^3), while significantly increasing the distant third contributor x^4 as well.

In conjunction with the minimum ridge results, Table 5.12 presents a tolerance limits analysis based on the *even* coded radius points. The tolerance limits succeed in covering the z_k response values that will occur on average at least 95% of the time; however, the skewed distribution in PGP2 can produce values for single instances of z_k considerably higher than indicated by either Z(**x**)

TABLE 5.9
EXPERIMENTAL DESIGN FOR PGP2

Coded x_K				*Response*
x^1	x^2	x^3	x^4	Z(**x**)
-1	-1	-1	-1	448.532
-1	-1	-1	+1	448.015
-1	-1	+1	-1	447.967
-1	-1	+1	+1	447.627
-1	+1	-1	-1	447.959
-1	+1	-1	+1	447.620
-1	+1	+1	-1	447.665
-1	+1	+1	+1	449.648
+1	-1	-1	-1	448.057
+1	-1	-1	+1	447.717
+1	-1	+1	-1	447.892
+1	-1	+1	+1	449.875
+1	+1	-1	-1	447.755
+1	+1	-1	+1	449.738
+1	+1	+1	-1	450.095
+1	+1	+1	+1	452.159
0	0	0	+3	450.566
0	0	0	-3	448.130
0	0	+6	0	455.702
0	0	-6	0	455.121
0	+6	0	0	454.375
0	-6	0	0	454.043
0	0	0	0	447.552

Uncoded Values				*Coded Value*
x^1	x^2	x^3	x^4	
2.271	4.605	5.045	5.567	0
2.071	4.405	4.845	5.367	-1
2.471	4.805	5.245	5.767	+1

or the tolerance limit's test for the maximum value (z_m). The scale of the probabilities $p(z_k)$ and their respective values z_k prevent graphing the probability distribution (e.g., Figure 4.6).

TABLE 5.10
REGRESSION RESULTS FOR CC DESIGN IN TABLE 5.9 FOR PGP2

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	R Square
Model	14	151.136	10.795	.9636
Error	8	5.715	.714	
Total	22			

Parameter Estimates

Variable	Coded Par. Est.	Uncod. Par. Est.
Intercept	447.552	1782.33
x^1	0.5159	-182.07
x^2	0.6102	-147.11
x^3	0.7515	-157.55
x^4	1.2163	-144.86
$x^1 \cdot x^1$	0.4900	12.25
$x^1 \cdot x^2$	2.0463	8.55
$x^2 \cdot x^2$	6.6563	4.62
$x^1 \cdot x^3$	2.2402	9.33
$x^2 \cdot x^3$	12.2771	8.53
$x^3 \cdot x^3$	7.8589	5.46
$x^1 \cdot x^4$	0.9195	7.66
$x^2 \cdot x^4$	5.5193	7.67
$x^3 \cdot x^4$	5.5160	7.66
$x^4 \cdot x^4$	1.7955	4.99

Based on these results, it appears that PGP2 represents a problem with a relatively flat surface in the region of optimality. Table 5.11 reports the coded radius' response found $Z(\mathbf{x}) = 447.805$ for the 1.0 coded radius along the minima ridge, while $Z(\mathbf{x}) = 459.002$ at the equivalent distance in the maxima direction. Furthermore, the tolerance analysis shows the worst case realization increasing along the minima ridge as well; consequently, unlike APL1P no tradeoff exists between increased $Z(\mathbf{x})$ and a more favorable underlying distribution.

TABLE 5.11
A CANONICAL ANALYSIS OF PGP2

	Eigenvectors			
Eigenvalues	x^1	x^2	x^3	x^4
14.7865	0.1102	0.6418	0.7017	0.2892
1.1515	-0.0018	0.6365	-0.7080	0.3059
0.5485	-0.0050	-0.4216	0.0127	0.9067
0.3143	0.9939	-0.0721	-0.0790	-0.0270

	Estimated Minima Ridge				
Coded Radius	x^1	x^2	x^3	x^4	Z(x)
0.0	2.271	4.605	5.045	5.567	447.55
0.2	2.254	4.641	5.048	5.461	447.62
0.4	2.233	4.716	5.074	5.364	447.65
0.6	2.209	4.794	5.096	5.275	447.71
0.8	2.181	4.871	5.117	5.194	447.76
1.0	2.150	4.945	5.137	5.124	447.80

	Estimated Maxima Ridge				
Coded Radius	x^1	x^2	x^3	x^4	Z(x)
0.0	2.271	4.605	5.045	5.567	447.55
0.2	2.277	4.750	5.206	5.616	449.54
0.4	2.282	4.903	5.374	5.652	451.77
0.6	2.287	5.057	5.542	5.687	454.04
0.8	2.291	5.211	5.711	5.722	456.47
1.0	2.295	5.365	5.879	5.757	459.00

TABLE 5.12
TOLERANCE LIMITS FOR PGP2 MINIMA RIDGE (RANDOM SEED = 34808)

			Population		Tol. Limit		
x^1, x^2, x^3, x^4 †	Z(x)	Mdn. z_k	Min z_k	Max z_k	Min z_k	Max z_k	% Cvg.
2.27, 4.61, 5.05, 5.67	447.55	466.84	185.07	8719.34	315.83	584.73	.9838
2.25, 4.64, 5.05, 5.46	447.62	466.36	184.56	8802.09	320.12	2250.94	.9900
2.23, 4.72, 5.07, 5.36	447.65	466.24	184.71	8818.87	334.65	575.31	.9706
2.21, 4.79, 5.10, 5.28	447.71	466.16	184.85	8830.72	321.92	558.81	.9683
2.18, 4.87, 5.12, 5.19	447.76	466.08	184.94	8842.56	335.18	702.33	.9804
2.12, 4.95, 5.14, 5.12	447.80	465.89	184.75	8879.04	339.01	761.46	.9547

† — Entries based on coded radius estimates in Table 5.11

PGP2 Analysis Summary. *Recommend a near-optimal solution of x^1 = 2.271, x^2 = 4.605, x^3 = 5.045, and x^4 = 5.567. If any adjustments in this solution must be made, then (1) avoid substantial increases in* ***both*** *x^2 and x^3; (2) try to significantly reduce x^4 and marginally reduce x^1 by proportionally increasing x^2 and x^3, respectively; and, (3) recognize that any change will likely increase the cost of the worst-case scenario.*

Finally, Table 5.13 presents the VRT results for PGP2 for selected factorial design points from the preliminary experimental design. With the exception of the second sample, which interestingly possesses a relatively high Z(**x**), the LH samples perform remarkably and consistently well. By contrast, the CV results vary considerably — from nearly 100% reduction to one case of variance *increase*. These results suggest that the LH method as the better VRT for the PGP2 problem. (The CV VRT uses ω^j, j = 1, 2, 3 for controls.)

TABLE 5.13
COMPARISON OF ESTIMATOR ACCURACY AND VARIANCE FOR RS, CV, AND LH SAMPLING TECHNIQUES FOR PGP2 (I=50, N=10)

x^1, x^2, x^3, x^4	Z(**x**)	$\bar{Z}_{RS}$	$\bar{s}^2_{RS}$	$\bar{Z}_{CV}$	$\bar{s}^2_{CV}$	%†	$\bar{Z}_{LH}$	$\bar{s}^2_{LH}$	%†
2.25, 4.55, 5.00, 5.50	447.7	452.67	101.1	450.19	113.8	-.13	445.45	6.1	.94
1.75, 4.05, 4.50, 5.00	489.5	525.75	1806.6	505.76	1673.9	.07	507.19	1301.1	.28
2.75, 5.05, 5.50, 6.00	458.8	458.83	95.2	458.68	0.1	1.00	459.65	2.6	.97
1.75, 4.05, 5.50, 6.00	447.9	454.05	222.6	448.65	71.1	.68	447.46	10.7	.95

† - % Variance reduction from RS; also note that '-' indicates variance increase.

5.3 CEP1

5.3.1 CEP1 Problem Description

CEP1 represents a two-stage machine capacity expansion problem donated to the University of Michigan FTP site by Higle and Sen (1990). The CEP1 recourse formulation possesses the same transportation problem structure seen in APL1P and PGP2, where again the first-stage variables model a capacity expansion decision for the supply nodes. However, CEP1 distinguishes itself from the previous problems in two ways: (1) the first-stage decision costs possess a piecewise linear structure, and (2) the first-stage variables have upper bounds. As shown in Figure 5.2, only variables x^5 through x^8 affect the recourse problem directly; however, each one's ability to do so beyond 500 depends upon the capacity decision associated with the variable pairings $\{x^j, x^{j+4}\}$, $j = 1,\ldots, 4$. In other words, for x^5 through x^8 the first 500 units are free; each additional unit above that point costs an amount associated with its paired variable. Consequently, the feasible region of CEP1 can be described in four-dimensional space $(x^5,\ldots,x^8)$ using a piecewise linear cost function; e.g., the cost of $x^{j+4} = c^j \cdot \text{MIN}[0, x^{j+4} - 500]$, $j = 1,\ldots, 4$. CEP1 also models a constraint using $x^5,\ldots, x^8$ where the upper bound is less than or equal to 100. The nature of this constraint is unknown to the author; hence, the text will refer to it as the 'joint' constraint.

The bounds on x^5 through x^8 also present a unique modification of the response surface analysis by further constraining the feasible region $\mathbf{Ax} = \mathbf{b}$ with $\mathbf{0} \leq \mathbf{x} \leq \mathbf{U_X}$ (where $\mathbf{U_X}$ represents their upper limits). In APL1P and PGP2, first-

$$
\begin{array}{lcccccccccc}
\text{MIN} & +2.5x^1 & +3.75x^2 & +5.0x^3 & +3.0x^4 & & & & & & \\
\text{S.T.} & -x^1 & & & & +x^5 & & & & & \leq 500 \\
 & & -x^2 & & & & +x^6 & & & & \leq 500 \\
 & & & -x^3 & & & & +x^7 & & & \leq 500 \\
 & & & & -x^4 & & & & +x^8 & & \leq 500 \\
 & & & & & +.08x^5 & +.04x^6 & +.03x^7 & +.01x^8 & & \leq 100 \\
 & & & & & +x^5 & & & & & \leq 2000 \\
 & & & & & & +x^6 & & & & \leq 2000 \\
 & & & & & & & +x^7 & & & \leq 3000 \\
 & & & & & & & & +x^8 & & \leq 3000 \\
 & & & & & -.8x^5 & & & & +\mathbf{W}^{4\bullet}\mathbf{y} & \leq 0 \\
 & & & & & & -x^6 & & & +\mathbf{W}^{5\bullet}\mathbf{y} & \leq 0 \\
 & & & & & & & -x^7 & & +\mathbf{W}^{6\bullet}\mathbf{y} & \leq 0 \\
 & & & & & & & & -x^8 & +\mathbf{W}^{7\bullet}\mathbf{y} & \leq 0
\end{array}
$$

Figure 5.2. CEP1 Formulation of First-Stage Variables
($\mathbf{W}^{i\bullet}$ Represents i^{th} Row of $\mathbf{W}$. Constraints Without $\mathbf{W}^{i\bullet}$ Comprise $\mathbf{Ax} \leq \mathbf{b}$)

stage decision variables *without* upper bounds allows for a response such that beginning with any non-optimal feasible $\mathbf{x}$ and proceeding in any descending search direction, $Z(\mathbf{x})$ will initially *decrease* due to the combined effects of (1) decreasing recourse costs disproportionally offsetting the increasing expense of those x^js *rising* in value; and, (2) reduced recourse *and* first-stage costs for *decreasing* x^js. This marginal cost reduction continues until reaching equilibrium at $Z(\mathbf{x}^*)$, after which the previous effects reverse themselves and drive $Z(\mathbf{x})$ back up — any additional supply of *increasing* x^j become surplus resources in the recourse problem, while any *decreasing* x^j cannot offset the additional cost of resource shortages. As the analysis will shortly show, the upper bounding in

CEP1 does not allow **x** to reach this equilibrium state between changes in **x** and the recourse costs, resulting in a truncated response surface in the $\{x^5,\ldots, x^8\}$ space.

Finally, like PGP2, CEP1 restricts its variation to the recourse right-side vector ω. The demand scenarios' sum total of 216 represents the lowest number of possible realizations of all the problems investigated by this dissertation. CEP1 also models surplus power availability to guarantee complete recourse for any value of **x** and realization of ω.

5.3.2 CEP1 Optimization Results

The OBS algorithm found 42 optimal bases and associated dual vectors for CEP1 (Table 5.14), with a frequency of occurrence listed in Table 5.15. In this particular problem the search techniques found six additional bases while using the OBS-COMPLETE method, which occurred primarily when $x^8 = 3000$ and one or more realizations of $\omega^j = 0$. This result implies that the undirected Monte Carlo search of the feasible space in the OBS algorithm did not sample this region adequately enough, and suggests that additional optimal bases may remain undetected. (This analysis of CEP1 uses the ODV method *after* the OBS-COMPLETE algorithm; thus, that algorithm's use of the expanded dual vector set ensures unbiased results.)

In a related matter, the OBS-COMPLETE technique turns in a better performance than the ODV method for all three optimal search techniques (Table 5.16), although again both outperform the OSL option by an order of magnitude.

TABLE 5.14
OBS-COMPLETE RESULTS FOR CEP1

Sample Size (ω - **Tx**)	Random # Seed	All Bases / First Optimal*	# Opt. Bases/ Dual Vectors	CPU Time (secs)
1000	18975	All	28	27.06
5000	913342061	All	33	55.48
50,000	159568	First	42	242.75
250,000	506886247	First	42†	120.88

* - 'First Optimal' option skips any remaining bases after finding first feasible, whereas 'All Bases' checks every basis in *P* for each sample (ω - **Tx**).
† - This run verified the third run using a larger sampling space. Subsequent search routines found 6 additional optimal bases.

TABLE 5.15
FREQUENCY OF BASIS OPTIMALITY FOR CEP1
(BASED ON 4th RUN FROM TABLE 5.14)

Basis ID #	1	2	3	4	5	6	7	8	9	10	11	12	13-42
Freq. of Optimality	.29	.13	.09	.09	.07	.04	.04	.03	.03	.03	.02	.02	.12
Cumulative Freq.*	.29	.42	.51	.60	.67	.71	.75	.78	.81	.84	.86	.88	1.00

* - May not add due to roundoff error.

TABLE 5.16
COMPUTATION TIMES OF OSL/OBS/ODV OPTIONS FOR GEOMETRIC SIMPLEX, PROJECTED GRADIENT, AND PARTAN ALGORITHMS FOR CEP1 (IN SECONDS)

Algorithm	OSL	OBS*	ODV
GEOMETRIC SIMPLEX (100 Iterations)	244.53	26.74	32.12
PROJECTED GRADIENT (6 Iterations)	96.87	6.89	8.03
PARTAN (7 Iterations)	261.62	18.90	21.43

* - Each search technique found two additional optimal bases.

This performance advantage probably results from the ODV algorithm having to check every array of a larger set of dual vectors (46), while the first 5 optimal bases provide a feasible answer two-thirds of the time (on average) for the OBS-

COMPLETE method. Consequently, given the marginal loss of coverage by the initial optimal basis set; the ability of OBS-COMPLETE to recognize infeasibility and supplement the basis set; and, its slight performance edge over the ODV method, the results of CEP1 support using the OBS-COMPLETE technique.

Regarding the search techniques, Tables 5.16 and 5.18 also show the PROJECTED GRADIENT ALGORITHM clearly outperforming both the GEOMETRIC SIMPLEX and PARTAN methods, again due to the small number of scenarios all search techniques calculate the true response $Z(\mathbf{x})$. The GEOMETRIC SIMPLEX ALGORITHM (Table 5.17) especially runs into difficulties with CEP1 due to the eventual collapse of the simplex into the expected value approximation $\mathbf{x}_{ev}$. Indeed, during the first 100 iterations the simplex finds an expansion move to replace the fifth vertex *every time*, and never initiates a shrinkage-enlargement-shrinkage cycle. This phenomenon occurs due to a combination of several factors: (1) the small coverage area of the collapsed simplex; (2) its location near the lower or upper bounds of two of the five variables (counting the slack); (3) sampling the entire population of responses z_{ij} due to the small number of scenarios; and, (4) the relatively steep slope of the response in this region of $\mathbf{x}$. Consequently, the simplex (lacking any real directional data) tends to incrementally move closer to the feasible boundaries using relatively smaller projection vectors; and, never cycles through the vertices since the steepness of the response and use of the true response $Z(\mathbf{x})$ (versus an estimate $\hat{Z}_s(\mathbf{x})$) almost guarantees a marginal improvement over the value of the next-worst vertex.

TABLE 5.17
SELECTED GEOMETRIC SIMPLEX MOVES FOR CEP1
(RANDOM SEED = 354644707)

k	x^5, x^6, x^7, x^8	$Z(\mathbf{x})^*$	Simplex Move	Replaces
0	0, 500, 1666, 3000	367014	Initial Vertex ($\mathbf{x}_{ev}$)	—
1	14, 1000, 823, 1172	867790	Expansion $\mathbf{x}_e$	5
2	28, 1142, 978, 1095	806841	Expansion $\mathbf{x}_e$	5
3	41, 718, 1344, 1530	704261	Expansion $\mathbf{x}_e$	5
4	10, 829, 1213, 1931	612543	Expansion $\mathbf{x}_e$	5
5	21, 767, 1371, 1995	573611	Expansion $\mathbf{x}_e$	5
10	2, 565, 1587, 2797	404966	Expansion $\mathbf{x}_e$	5
20	0, 500, 1665, 2995	368060	Expansion $\mathbf{x}_e$	5
40	0, 500, 1666, 3000	367014	Expansion $\mathbf{x}_e$	5
60	0, 500, 1666, 3000	366972	Expansion $\mathbf{x}_e$	5
100	0, 500, 1666, 3000	366951	Expansion $\mathbf{x}_e$	5

* - Differences in $Z(\mathbf{x})$ due to fractional components of $\mathbf{x}$.

By contrast the PROJECTED GRADIENT ALGORITHM (Table 5.18) performs very well, although it requires inputs of 1.0 for the scalar multiple of the projection vector in every case. The behavior of this algorithm provides the clearest evidence of the truncated nature of the response surface for CEP1 in the following ways.

1. *Quadratic Estimates.* The quadratic regression of the projection vector fits extremely well at every iteration; furthermore, the scalar multiple exceeds 1.0 in each case as well. This implies that the equilibrium point remains well below the current optimal solution.

TABLE 5.18
PROJECTED GRADIENT ITERATIONS FOR CEP1
(SCENARIOS = 216, $Q = 8$)

k	x^5, x^6, x^7, x^8	$\mathbf{d}^k$	Est. q	Act. q	R^2	Z(x)
1	500, 500, 500, 500	-.23, .33, .36, .49	5.2	1.0	.999	1,234,278
2	31, 1155, 1218, 1478	-.31, .26, .33, .49	19.9	1.0	.999	640,930
3	0, 1181, 1250, 1526	0, -.13, .04, .39	3.5	1.0	.999	618,832
4	0, 693, 1409, 3000	0, -.05, .06, 0	1.4	1.0	.999	374,482
5	0, 0, 2333, 3000	0, 0, -.12, .35	1.3	1.0	1.00	355,160
6	0, 0, 2333, 3000*	0, 0, 0, 0	—	—	—	355,160

* - Optimal solution.

2. *Descent Gradient.* Table 5.18 reports the normalized *projected* descent gradient $\mathbf{d}_6 = 0$, thus implying an optimal solution since $\mathbf{d}_k$ is derived from the true unconstrained descent gradient $-\nabla Z(\mathbf{x}_6)$. Given the non-differentiable property of $E[h(\mathbf{x},\omega,\mathbf{T})]$, such a condition can occur either through multiple optimality or binding constraints as expressed in the working set. Furthermore, the unprojected descent gradient $-\nabla Z(\mathbf{x}_k)$ remains relatively large (-83, -189, -155, and -170 for x^5 through x^8, respectively) at the optimal solution. This indicates that an unconstrained environment would allow further reductions in $Z(\mathbf{x})$ from the current position.

Unlike the GEOMETRIC SIMPLEX ALGORITHM, the PROJECTED GRADIENT ALGORITHM's strong directional capabilities, coupled with scalar estimates enhanced by the lack of a flat region, finds the optimal solution quickly.

The PARTAN ALGORITHM also performs well, although it suffers somewhat from the presence of upper bounds on the first-stage decision variables (Table 5.19). Recalling that the PARTAN method derives its directional guidance from an alternating gradient and line search approach, PARTAN momentarily stalls when $\mathbf{x}_{k-1}$ and $\mathbf{p}_k$ approach the same point (0, 693, 1409, 3000). However, unlike the GEOMETRIC SIMPLEX ALGORITHM, PARTAN regains its bearings due to its gradient capabilities, and proceeds to the optimal solution relatively quickly. Unfortunately, the assumed advantage of the PARTAN approach — theoretical convergence in n-1 iterations for a quadratic function of n parameters — is not realized in this case. Indeed, any advantages PARTAN has in this instance comes from its projected gradient component, which the PROJECTED GRADIENT ALGORITHM itself provides more directly with better results.

Finally, it should be noted that using the optimal basis or dual vector sets can give slightly different directional descent information than OSL due to the likelihood of multiple optimality for certain values of $\mathbf{x}_k$, which in turn can affect the convergence rate of the search. The most notable example of this phenomenon occurs with the PARTAN ALGORITHM, where the OBS-COMPLETE version takes two additional iterations at the end to confirm optimality. Additionally, the estimates of the quadratic fit vary considerably between the OSL, OBS-COMPLETE, and ODV versions on the last iteration. This variation occurs since a combination of short distances and slightly different projection vectors produce different versions of near-optimal sample values of $Z(\mathbf{x})$.

TABLE 5.19
PARTAN ITERATIONS FOR CEP1 (OSL AND ODV) (SCENARIOS = 216, $Q = 8$)

k	x^5, x^6, x^7, x^8	Est. q^*	Act. q	R^2	Z(x)
$\mathbf{x}_0$	500, 500, 500, 500	4.1	1.0	.999	1,234,278
$\mathbf{x}_1$	31, 1155, 1218, 1478	4.2	1.0	.999	640,930
$\mathbf{p}_1$	0, 1181, 1250, 1526	3.9	1.0	.999	618,832
$\mathbf{x}_2$	0, 1181, 1250, 1526	2.3	1.0	.999	618,832
$\mathbf{p}_2$	0, 693, 1409, 3000	2.3	1.0	.999	374,482
$\mathbf{x}_3$	0, 693, 1409, 3000	2.3	1.0	.999	374,482
$\mathbf{p}_3$	0, 693, 1409, 3000	2.3	1.0	.999	374,482
$\mathbf{x}_4$	0, 693, 1409, 3000	2.3	1.0	.999	374,482
$\mathbf{p}_4$	0, 693, 1409, 3000	2.4	1.0	.999	374,482
$\mathbf{x}_5$	0, 693, 1409, 3000	1.1	1.0	.999	374,482
$\mathbf{p}_5$	0, 0, 2333, 3000	1.2	1.0	.999	355,160
$\mathbf{x}_6$	0, 0, 2333, 3000	2.8	1.0	.999	355,160
$\mathbf{p}_6$	0, 0, 2333, 3000	1.1	1.0	.999	355,160
$\mathbf{x}_7$	0, 0, 2333, 3000	—	—	—	355,160

5.3.3 CEP1 Response Surface Analysis

The statistical analysis of CEP1 becomes somewhat abbreviated due to the nature of the optimal solution imposed by the upper bounds on the decision variables. The fundamental assumption made by this analysis — the existence of multiple-optimal or near-optimal solutions within a 'flat' region — obviously does not occur in this case. Instead, the optimal solution lies on the 'side' of the unconstrained region, prevented from moving towards the equilibrium point at the 'bottom' by the bounds imposed on the decision variables. Therefore, *any* movement in *any* feasible direction away from the optimal solution causes a steep

increase in Z(**x**). By contrast, any relaxation of the upper bound constraints on x^7 or x^8 produces a considerable decrease in Z(**x**).

However, this result does not preclude describing the distributional nature of the current optimal solution using tolerance limits. Indeed, an analysis of several feasible solutions reveals a very skewed frequency distribution as indicated by the large discrepancies between Z(**x**) and the median (Table 5.20). Figure 5.3 on the following page reinforces this description with a graphical portrait of the population distribution for the optimal solution. Although Z(**x**) remains the best measure of the long-term operating costs, the tolerance limit analysis clearly shows that occurrences requiring much higher expenditures will very likely occur over the lifetime of this problem.

Although a formal response surface approximation as originally intended for these type of problems cannot be performed for CEP1, the insight obtained can still be presented to the decision-maker.

TABLE 5.20
TOLERANCE LIMITS FOR CEP1 (RANDOM SEED = 221789)

			Population		Tol. Limit		
x^5, x^6, x^7, x^8	Z(**x**)	Mdn. z_k	Min z_k	Max z_k	Min z_k	Max z_k	% Cvg.
0, 0, 2333.33, 3000	355,160	154,013	16,667	1,833,413	18,542	1,593,938	.9907
0, 0, 2000, 3000	408,826	269,775	15,000	1,950,350	17,000	1,710,350	.9815
0, 0, 2500, 2500	419,729	252,432	16,000	1,931,844	17,875	1,691,844	.9861
0, 0, 2666.66, 2000	493,034	350,800	15,333	2,030,275	15,333	1,790,800	.9954

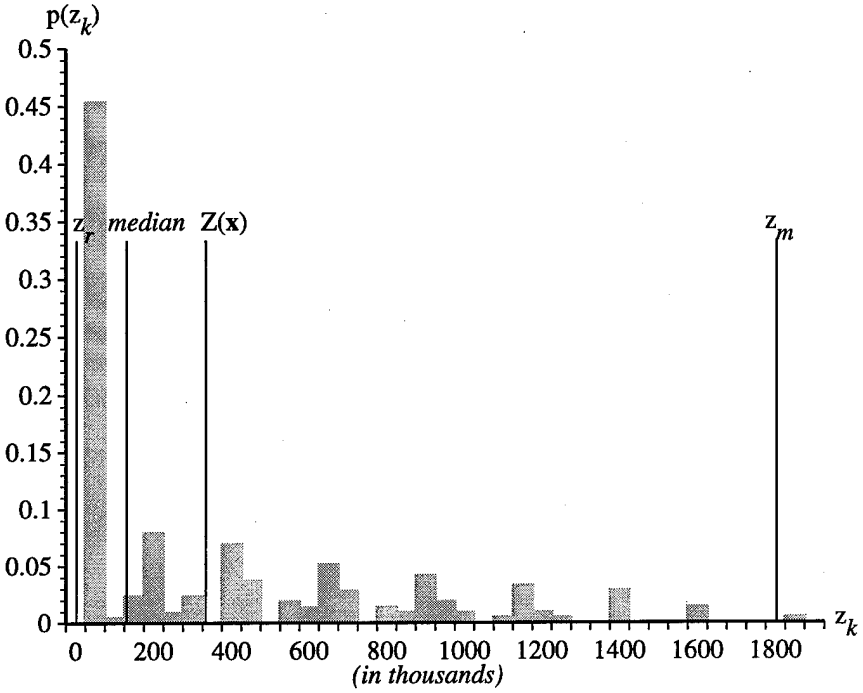


Figure 5.3. Comparison of Tolerance Limits to Population Distribution, Mean, and Median for $\mathbf{x}_k = (0, 0, 2333.33, 3000)$

CEP1 Analysis Summary. *The optimal solution $x^7 = 2333.33$ and $x^8 = 3000$ can be substantially improved by **relaxing the upper bounds** on either x^8 or the 'joint' constraint. Any feasible deviation from the optimal solution under current constraints will considerably **increase** $Z(\mathbf{x})$ and most likely raise the cost of the worst case scenario. Finally, scenarios costing **four times higher** than average are possible.*

Both VRT methods significantly reduce the variance of the estimators of $Z(\mathbf{x})$ in the case of CEP1. Table 5.21 shows the results where, unlike PGP2, the

CV technique remains competitive with the LH approach (CVs use all three ω^j as controls). In both PGP2 and CEP1 the large skewed characteristic of the underlying distribution most likely accounts for reducing VRT effectiveness when compared to APL1P. Nonetheless, both techniques offer considerable improvement over the random sample estimator.

TABLE 5.21
COMPARISON OF ESTIMATOR ACCURACY AND VARIANCE FOR RS, CV, AND LH SAMPLING TECHNIQUES FOR CEP1 (I=50, N=10)

x^5, x^6, x^7, x^8	Z(x)	$\bar{Z}_{RS}$	$\bar{s}^2_{RS}$	$\bar{Z}_{CV}$	$\bar{s}^2_{CV}$*	%†	$\bar{Z}_{LH}$	$\bar{s}^2_{LH}$*	%†
0, 0, 2333, 3000	355160	361058	3.212	354411	.519	.84	346827	.365	.89
0, 0, 2000, 3000	408826	406321	5.714	397896	.784	.86	407738	.570	.90
0, 0, 2500, 2500	419729	415518	1.689	419349	.431	.75	412513	1.02	.39
0, 0, 2666, 2000	493034	522554	2.180	497606	.203	.91	513481	.287	.87

* - in billions. † - % Variance reduction from RS

5.4 4TERM

5.4.1 4TERM Problem Description

4TERM models a vehicle allocation problem between a central depot and four outlying terminals. The vehicles are single tractors with a one- or two-trailer configuration, while demand constitutes the stochastic right-side elements modeling daily pick-up and delivery requirements at each of the four terminals (for a total of eight independent right-side random variables). Each random variable can take one of two discrete values with equal probability, thus providing 256 possible demand scenarios. The first-stage decision variables model the

allocation — or basing — of the existing fleet of 300 trailers and 200 tractors among the five locations using the constraints

$$\sum_{j=1}^{5} x^j = 300 \ \textit{(Trailers)} \tag{5.1a}$$

$$\sum_{j=6}^{10} x^j = 200 \ \textit{(Tractors)} \tag{5.1b}$$

while the availability of single tractor-trailer combinations for daily rent to supplement the existing fleet is represented as

$$\sum_{j=11}^{15} x^j \leq 10{,}000 \ \textit{(Rental Tractor-Trailer)}. \tag{5.2}$$

The cost of using $x^j, j = 11, \ldots, 15$ is $c^j = 100$, while the existing fleet's expenses are zero under this model; i.e., $c^j = 0, j = 1, \ldots, 10$. Equations (5.1) and (5.2) constitute $\mathbf{Ax} = \mathbf{b}$, and $\mathbf{x} \geq \mathbf{0}$.

The **T** matrix deterministically allocates the decision variables among the four terminals and central depot without any gains, losses, or stochastic representation; e.g., x^1 (trailers) and x^6 (tractors) model the transport resource availability of the central depot. A single recourse right-side variable models each type of resource — tractor or trailer — separately; thus, $x^j, j = 1,\ldots, 10$ correspond directly to their own resource element. The rental decision vector also transfers to the recourse right-side under the same conditions, with the exception that as a tractor-trailer package it adds resources to *two* separate elements; e.g., x^{11} adds one trailer and one tractor to the central depot resources supplied by x^1 and x^6, respectively. Mathematically, the **T** matrix can be expressed as

$$-x^j - x^{j+10} = 0,\ j = 1,\ldots, 5 \tag{5.3a}$$

$$-x^j - x^{j+5} = 0,\ j = 6,\ldots, 10. \tag{5.3b}$$

The recourse model attempts to retain the same home-base allocation represented by the first-stage vector $\mathbf{x}$; however, penalty costs of 1,000 per tractor or trailer allow for complete recourse if there is an insufficient number of vehicles or if mismatched demand occurs.

5.4.2 4TERM Optimization Results

Unlike the previous problems, 4TERM does not lend itself to a manageable number of optimal bases for the OBS-COMPLETE or ODV algorithms; consequently, OBS-RESET remains the only computational alternative to OSL for calculating $Z(\mathbf{x})$. In this case, the algorithm resets the optimal basis and dual vector sets to zero for each individual $\mathbf{x}_k$, which in turn provides a noticeable reduction in computation times. However, as Table 5.22 shows, the amount of reduction does not approach the OBS-COMPLETE/ODV results of the other problems (assuming its availability). Although several $\mathbf{x}_k$ require a unique basis to cover each of the possible 256 scenarios, most only need between 50 and 60.

Table 5.23 presents the GEOMETRIC SIMPLEX ALGORITHM results for 4TERM. Again due to the small number of scenarios, all the search methods find the true response $Z(\mathbf{x})$. Unlike previous problems the simplex avoids a

TABLE 5.22
COMPUTATION TIMES OF SELECTED OPTIONS FOR GEOMETRIC SIMPLEX, PROJECTED GRADIENT, AND PARTAN ALGORITHMS FOR 4TERM (IN SECONDS)

Algorithm	OSL	OBS-RESET
GEOMETRIC SIMPLEX (100 Iterations)	848.59	262.52
PROJECTED GRADIENT (13 Iterations)	125.39	91.19
PARTAN (27 Iterations)	591.60	449.11

premature collapse into the vertex $\mathbf{x}_{ev}$; indeed, the terminating vertex $\mathbf{x}_{100}$'s objective function value of 42369 represents a substantial improvement over $Z(\mathbf{x}_{ev})$. However, comparing the vertices of the last simplex indicates that a slow convergence begins in the latter stages of the search similar to what occurs with PGP2 and CEP1. Furthermore, the simplex's computational time already approaches an order of magnitude higher (OSL option) than the PROJECTED GRADIENT ALGORITHM's while still somewhat far from a near-optimal solution. This persistent tendency of the simplex to prematurely converge casts doubt on its ability to continue towards the region of optimality in a reasonable amount of time. Finally, the GEOMETRIC SIMPLEX ALGORITHM tends to drop the $x^{11},\ldots, x^{15}$ values at a steady pace. As seen shortly, the PROJECTED GRADIENT ALGORITHM initiates a similar steep drop in these decision variables, thus validating the simplex's sensitivity to the constraint (5.2). This inclination of x^{11} through x^{15} to drop to zero early will essentially eliminate them from the experimental design of 4TERM.

TABLE 5.23
SELECTED GEOMETRIC SIMPLEX MOVES FOR 4TERM
(RANDOM SEED = 800598600)

k	$\mathbf{x}_k$	Z(x)	Simplex Move	Replaces
0	192, 20, 50, 8, 30 146, 10, 25, 4, 150 0, 0, 0, 0, 0	53313	Initial Vertex ($\mathbf{x}_{ev}$)	—
1	71, 51, 110, 35, 33 75, 38, 22, 17, 47 904, 1881, 1631, 398, 1829	699702	Expansion $\mathbf{x}_e$	16
10	85, 28, 128, 23, 36 83, 41, 17, 12, 46 879, 881, 612, 1025, 1216	496764	Expansion $\mathbf{x}_e$	16
20	118, 35, 80, 11, 56 102, 23, 16, 13, 46 530, 414, 545, 506, 614	296387	Expansion $\mathbf{x}_e$	16
30	162, 25, 62, 17, 35 123, 14, 16, 18, 30 512, 311, 310, 375, 326	218909	Expansion $\mathbf{x}_e$	16
40	186, 20, 46, 11, 38 124, 8, 15, 20, 33 300, 113, 171, 118, 203	125958	Expansion $\mathbf{x}_e$	16
50	199, 20, 39, 10, 32 129, 8, 15, 21, 27 287, 105, 85, 84, 101	101767	Expansion $\mathbf{x}_e$	16
60	201, 18, 40, 11, 30 134, 7, 17, 18, 24 228, 62, 29, 41, 93	80950	Expansion $\mathbf{x}_e$	16
80	193, 20, 47, 8, 31 141, 10, 23, 8, 19 50, 29, 16, 18, 29	49609	Expansion $\mathbf{x}_e$	16
90	193, 20, 48, 8, 31 142, 9, 23, 7, 18 37, 19, 20, 14, 17	46221	Expansion $\mathbf{x}_e$	16
100	192, 20, 49, 8, 31 143, 10, 24, 6, 17 15, 13, 10, 13, 14	42369	Expansion $\mathbf{x}_e$	16

Fortunately, as in previous problems the PROJECTED GRADIENT ALGORITHM performs extremely well, reaching an optimal solution in 13 iterations. Indeed, as Table 5.24 shows on the following page, the algorithm descends very rapidly towards the region of optimality in the first seven iterations before taking another six steps to the optimal solution. Unlike APL1P and PGP2, the unprojected directional descent vector at optimality is zero, thus implying a 'flat' region composed of multiple optimal solutions. Subsequent searches — specifically PARTAN and those conducted for the experimental design centerpoint explained shortly — confirmed multiple optimality by finding different solutions in a similar number of steps. Furthermore, the PROJECTED GRADIENT ALGORITHM, like all search algorithms this research applies to 4TERM, immediately begins eliminating the rental tractor-trailer variables ($x^{11},\ldots, x^{15}$). Typically, these variables drop to zero well before $Z(\mathbf{x})$ approaches 100,000, thus providing the basis for eliminating $x^{11},\ldots, x^{15}$ from the final experimental design. In effect, the PROJECTED GRADIENT ALGORITHM not only finds the optimal solution, but performs a factor screening function as well.

As in the previous problems, the PARTAN ALGORITHM (Table 5.25) converges to an optimal solution as well. Following the same pattern as before, the parallel tangent property does not appear to supply any additional advantages over using the projected gradient portion alone. However, starting from the same initial point as the PROJECTED GRADIENT ALGORITHM, the PARTAN ALGORITHM finds a different optimal solution, thus confirming multiple optimality of 4TERM.

TABLE 5.24
SELECTED PROJECTED GRADIENT ITERATIONS FOR 4TERM
(SCENARIOS = 256, $Q = 8$)†

k	$x^1 \ldots x^5$ $x^6 \ldots x^{10}$ $x^{11} \ldots x^{15}$	$d^1 \ldots d^5$ $d^6 \ldots d^{10}$ $d^{11} \ldots d^{15}$	Est. q	Act. q	R^2	Z(x)
1	60, … , 60 40, … , 40 2*k*, … , 2*k*	0, … , 0 0, … , 0 -.07, -.07, -.07, -.07, -.07	6.8	1.0	1.0	1035*k*
2	60, … , 60 40, … , 40 0, … , 0	.58, -.15, -.15, -.15, -.15 0, … , 0 0, … , 0	.31	.31	.99	73795
3	135, 41, 41, 41, 41 40, … , 40 0, … , 0	-.01, -.01, .03, -.01, -.01 .58, -.14, -.14, -.14, -.14 0, … , 0	.31	.31	.98	52535
4	134, 41, 44, 41, 41 90, 28, 28, 28, 28 0, … , 0	-.15, -.15, .59, -.15, -.15 0, … , 0 0, … , 0	.13	.13	.99	41096
5	129, 35, 66, 35, 35 90, 28, 28, 28, 28 0, … , 0	-.07, -.07, -.07, -.07, .28 -.14, -.14, .54, -.14, -.14 0, … , 0	.23	.23	.99	38930
6	125, 32, 62, 32, 48 83, 21, 53, 21, 21 0, … , 0	.17, -.04, -.04, -.04, -.04 0, … , 0 0, … , 0	.05	.05	.99	35723
7	132, 30, 61, 30, 47 83, 21, 53, 21, 21 0, … , 0	.00, -.00, -.00, -.00, -.00 0, … , 0 0, … , 0	.01	.01	.99	35518
13	137, 32, 64, 21, 47* 83, 21, 53, 21, 21 0, … , 0	0, … , 0 0, … , 0 0, … , 0	—	—	—	35514

* - Optimal solution. † - *k* represents 1,000.

TABLE 5.25
SELECTED PARTAN ITERATIONS FOR 4TERM (SCENARIOS = 256, $Q = 8$)

k	$x^1 \ldots x^5$ $x^6 \ldots x^{10}$ $x^{11} \ldots x^{15}$	Est. q^*	Act. q	R^2	Z(x)
$\mathbf{x}_0$	60, ... , 60 40, ... , 40 $2k$, ... , $2k$	6.75	1.00	1.0	1035014
$\mathbf{x}_1$	60, ... , 60 40, ... , 40 0, ... , 0	.312	.21	.99	73795
$\mathbf{p}_1$	111, 47, 47, 47, 47 40, ... , 40 0, ... , 0	15.18	1.00	.99	52033
$\mathbf{p}_2$	111, 46, 50, 46, 46 89, 28, 28, 28, 28 0, ... , 0	.361	.361	.99	39481
$\mathbf{p}_3$	117, 41, 60, 41, 41 98, 26, 26, 26, 26 0, ... , 0	.306	.306	.98	38864
$\mathbf{x}_7$	118, 40, 61, 40, 40 75, 18, 58, 18, 32 0, ... , 0	.06	.06	.99	36131
$\mathbf{p}_{18}$	128, 30, 65, 30, 47 75, 18, 59, 18, 30 0, ... , 0	.06	.01	.99	35611
$\mathbf{p}_{23}$	133, 33, 62, 27, 45 75, 18, 59, 18, 30 0, ... , 0	.13	.13	.99	35517
$\mathbf{x}_{26}$	139, 32, 60, 26, 43 75, 18, 59, 18, 30 0, ... , 0	—	—	—	35514

5.4.3 4TERM Response Surface Analysis

Several aspects of 4TERM pose additional challenges to fitting a polynomial approximation to $Z(\mathbf{x})$; specifically, the dimension of the decision vector $\mathbf{x}$, the tractor/trailer equality constraints (5.1), and its multiple optimal solutions. As indicated already, the gradient search techniques provide strong evidence that x^{11} through x^{15} simply do not assume a role in the region of optimal or near-optimal solutions. If any doubt persists regarding such an assumption, a factor screening design could help verify these results. However, this research considers the evidence from the gradient methods strong enough for the response analysis to proceed *without* considering the rental tractor-trailer decision variables.

Eliminating these five variables reduces the remaining number of variables under consideration to ten (x^1 through x^{10}). However, the equality constraints (5.1) do not allow the necessary degrees of freedom to construct a 10-variable central composite design; indeed, two variables — one each from the tractor and trailer groupings — must be 'thrown out' in order to construct a CCD design on the remaining eight. In effect, this reduction projects the true response onto the hyperplane defined by the remaining variables, thus providing a framework for applying 'standard' designs. This in turn suggests making *a priori* judgments on which projection would undergo the least distortion, and could consider such items as constraint coefficients, solution comparisons, or subjective interests. In 4TERM, this research eliminates x^1 and x^6 due to (1) their large values; (2) subjective interest in the outlying terminals (x^1 and x^6 represent the *central*

terminal's trailer and tractor capacity, respectively); and, (3) their *relatively* small variation among optimal solutions. (Other alternatives not explored by this research for handling input factor linear dependence include simplex-lattice designs (Scheffe 1958) and special linear transformations (Draper and Lawrence 1965a,b; Thompson and Myers 1968). Cornell (1973, 1979, 1981) also reviews these approaches with an accompanying bibliography.)

Multiple optimal solutions pose another challenge to the accuracy of the response surface approximation. Preliminary experimental designs varying in size, but all using the PROJECTED GRADIENT ALGORITHM's optimal solution, gave polynomial approximations with either positive indefiniteness (one of the eight eigenvalues slightly negative) or inadequate fits (R^2 values below .8). Since the presence of multiple optimality implies that the centerpoint can lie anywhere in a flat region, these results suggest the optimal solution from the PROJECTED GRADIENT ALGORITHM lies near the 'edge' of the region, and that a better fit can be derived using a more centrally located optimal solution. Such a centerpoint can be found using a convex combination of additional optimal solutions discovered by the PROJECTED GRADIENT ALGORITHM under different initial starting points. Table 5.26 presents five such optimal solutions; since $\mathbf{x}_3$ and $\mathbf{x}_5$ represent the two solutions furthermost apart ($\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are fairly close), the design centerpoint $\mathbf{x}_c$ represents an average of those two extremes.

The final experimental design (Table 5.27) employs 64 runs in the CCD portion (a quarter of the 256 full factorial), 16 axial points with coded multipliers of 2, and one centerpoint. The design possesses a resolution level of V, which

TABLE 5.26
DERIVATION OF EXPERIMENTAL DESIGN CENTERPOINT $\mathbf{x}^*_c$

$\mathbf{x}^*_k$	x^1	x^2	x^3	x^4	x^5	x^6	x^7	x^8	x^9	x^{10}	α_k
$\mathbf{x}^*_1$	137	32	64	21	46	84	21	53	21	21	.0
$\mathbf{x}^*_2$	138	39	66	15	42	69	41	30	40	20	.0
$\mathbf{x}^*_3$	151	32	60	15	42	78	19	35	19	49	.5
$\mathbf{x}^*_4$	137	30	61	15	57	87	17	40	29	27	.0
$\mathbf{x}^*_5$	137	47	61	15	40	108	24	35	8	25	.5
$\mathbf{x}^*_c$	144	39.5	60.5	15	41	93	21.5	35	13.5	37	

insures no two-way interaction confounding, and follows a design generator from Lorenzen and Anderson (1993). A full-factorial design for six variables — x^2, x^3, x^4, x^7, x^8, and x^9 — provides the structure for the $2^6 = 64$ CCD; then x^5's coded value is set by multiplying the coded values of x^2, x^4, x^8, and x^9, while multiplying x^3, x^4, x^7, x^8, and x^9 determines x^{10}.

Table 5.28 shows the polynomial approximation based on the data in Table 5.27, while Table 5.29 reports the eigenvalue, eigenvector, and ridge results. The regression supplies an acceptable fit (R^2 of .90) with all eigenvalues positive (positive definite fit), thus assuring a reasonable approximation of the projected response. As before, the best information comes from the canonical analysis, which in 4TERM's case provides several very useful insights. First of all, variables x^5 and x^8 dominate the two axes most sensitive to changes in the rotated coordinate system, while x^7, x^9, and x^{10} each heavily contribute, respectively, to the three least sensitive axes. Consequently, when examining the estimated ridge of maximum response, x^5 and x^8 drop significantly, with little

TABLE 5.27
EXPERIMENTAL DESIGN FOR 4TERM

Coded x_k								*Response*
x^2	x^3	x^4	x^5	x^7	x^8	x^9	x^{10}	Z(x)
-1	-1	-1	1	-1	-1	-1	-1	41206.921875
1	-1	-1	-1	-1	-1	-1	-1	45604.984375
-1	1	-1	1	-1	-1	-1	1	41206.921875
1	1	-1	-1	-1	-1	-1	1	45604.984375
-1	-1	-1	1	1	-1	-1	1	41206.921875
1	-1	-1	-1	1	-1	-1	1	45604.984375
-1	1	-1	1	1	-1	-1	-1	41206.921875
1	1	-1	-1	1	-1	-1	-1	45604.984375
-1	-1	-1	-1	-1	1	-1	1	45252.585938
1	-1	-1	1	-1	1	-1	1	40854.507812
-1	1	-1	-1	-1	1	-1	-1	42080.984375
1	1	-1	1	-1	1	-1	-1	38189.906250
-1	-1	-1	-1	1	1	-1	-1	45252.585938
1	-1	-1	1	1	1	-1	-1	40854.507812
-1	1	-1	-1	1	1	-1	1	42080.984375
1	1	-1	1	1	1	-1	1	38128.679688
-1	-1	1	-1	-1	-1	-1	1	43439.085938
1	-1	1	1	-1	-1	-1	1	39038.375000
-1	1	1	-1	-1	-1	-1	-1	43439.085938
1	1	1	1	-1	-1	-1	-1	39733.085938
-1	-1	1	-1	1	-1	-1	-1	43439.078125
1	-1	1	1	1	-1	-1	-1	39038.375000
-1	1	1	-1	1	-1	-1	1	43439.085938
1	1	1	1	1	-1	-1	1	39703.816406
-1	-1	1	1	-1	1	-1	-1	38685.945312
1	-1	1	-1	-1	1	-1	-1	43086.664062
-1	1	1	1	-1	1	-1	1	35575.906250
1	1	1	-1	-1	1	-1	1	39915.078125
-1	-1	1	1	1	1	-1	1	38685.945312
1	-1	1	-1	1	1	-1	1	43086.664062
-1	1	1	1	1	1	-1	-1	35575.906250
1	1	1	-1	1	1	-1	-1	39915.078125
-1	-1	-1	-1	-1	-1	1	1	45604.984375
1	-1	-1	1	-1	-1	1	1	41206.921875
-1	1	-1	-1	-1	-1	1	-1	45604.984375
1	1	-1	1	-1	-1	1	-1	41230.796875
-1	-1	-1	-1	1	-1	1	-1	45604.984375
1	-1	-1	1	1	-1	1	-1	41206.921875
-1	1	-1	-1	1	-1	1	1	45604.984375
1	1	-1	1	1	-1	1	1	41217.343750
-1	-1	-1	1	-1	1	1	-1	40854.507812
1	-1	-1	-1	-1	1	1	-1	45252.585938
-1	1	-1	1	-1	1	1	1	37682.906250
1	1	-1	-1	-1	1	1	1	42080.984375
-1	-1	-1	1	1	1	1	1	40854.507812

TABLE 5.27 — CONTINUED

Coded x_k								*Response*
x^2	x^3	x^4	x^5	x^7	x^8	x^9	x^{10}	Z(x)
1	-1	-1	-1	1	1	1	1	45252.585938
-1	1	-1	1	1	1	1	-1	37682.906250
1	1	-1	-1	1	1	1	-1	42080.984375
-1	-1	1	1	-1	-1	1	-1	39038.375000
1	-1	1	-1	-1	-1	1	-1	43439.085938
-1	1	1	1	-1	-1	1	1	39038.375000
1	1	1	-1	-1	-1	1	1	43439.078125
-1	-1	1	1	1	-1	1	1	39038.375000
1	-1	1	-1	1	-1	1	1	43439.078125
-1	1	1	1	1	-1	1	-1	39038.375000
1	1	1	-1	1	-1	1	-1	43439.093750
-1	-1	1	-1	-1	1	1	1	43086.671875
1	-1	1	1	-1	1	1	1	38685.945312
-1	1	1	-1	-1	1	1	-1	39915.078125
1	1	1	1	-1	1	1	-1	37220.449219
-1	-1	1	-1	1	1	1	-1	43086.664062
1	-1	1	1	1	1	1	-1	38685.937500
-1	1	1	-1	1	1	1	1	40003.546875
1	1	1	1	1	1	1	1	37186.652344
2	0	0	0	0	0	0	0	35801.640625
-2	0	0	0	0	0	0	0	41473.578125
0	2	0	0	0	0	0	0	35802.417969
0	-2	0	0	0	0	0	0	42209.976562
0	0	2	0	0	0	0	0	35516.523438
0	0	-2	0	0	0	0	0	42065.859375
0	0	0	2	0	0	0	0	35802.417969
0	0	0	-2	0	0	0	0	47621.117188
0	0	0	0	0	0	0	2	35514.363281
0	0	0	0	0	0	0	-2	37666.898438
0	0	0	0	2	0	0	0	35514.363281
0	0	0	0	-2	0	0	0	38013.695312
0	0	0	0	0	2	0	0	35514.363281
0	0	0	0	0	-2	0	0	49649.039062
0	0	0	0	0	0	2	0	35514.363281
0	0	0	0	0	0	-2	0	38764.906250
0	0	0	0	0	0	0	0	35514.363281

Uncoded Values								*Coded Values*
x^2	x^3	x^4	x^5	x^7	x^8	x^9	x^{10}	
40	61	15	40	22	35	13	37	0
30	51	5	30	17	25	8	27	-1
50	71	25	50	27	45	18	47	1

TABLE 5.28
REGRESSION RESULTS FOR CC DESIGN IN TABLE 5.27 FOR 4TERM

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	R Square
Model	44	807372056	18349365	.9001
Error	36	89585747	2488493	
Total	80	896957803		

Selected Parameter Estimates

Variable	Coded Par. Est.	Uncod. Par. Est.
Intercept	36182	141783
x^3	-1601	-748
x^4	-2180	-1183
x^5	-4407	-1435
x^8	-2421	-830
$x^2 \cdot x^2$	2372	6
$x^3 \cdot x^3$	2740	7
$x^3 \cdot x^8$	-2975	-7
$x^4 \cdot x^4$	2525	25
$x^5 \cdot x^5$	5446	14
$x^7 \cdot x^7$	498	5
$x^8 \cdot x^8$	6316	16
$x^9 \cdot x^9$	874	9

change in the remaining variables (of course, the eliminated variables x^1 and x^6 make up the difference).

Designating the outlying terminals *A* (where x^2 and x^7 represent the basing of terminal *A*'s trailers and tractors, respectively), *B* (x^3,x^8), *C* (x^4,x^9), and *D* (x^5,x^{10}), these results show that minimal increases from the optimal solution represented by the centerpoint involve significant reallocation of (1) trailer resources from the central terminal to the outlying nodes *B*, *C*, and *D*, and (2) tractors from the central terminal to nodes *B* and *D*. By contrast, the practical insight of the maxima ridge strongly suggests ***not reducing*** the number of trailers

TABLE 5.29
A CANONICAL ANALYSIS OF 4TERM

		Eigenvectors		
Eigenvalues	x^2	x^3	x^4	x^5
6856	.0058	-.3387	.0002	.0198
5476	.0626	.0696	.0450	.9946
2660	.5040	.3731	.7594	-.0932
2346	.7075	.2919	-.6344	-.0366
2072	-.4913	.8091	-.1274	-.0212
864	-.0134	-.0374	-.0499	-.0021
609	.0058	.0011	-.0017	.0019
214	.0011	.0001	-.0004	.0004

Eigenvalues	x^7	x^8	x^9	x^{10}
6856	.0002	.9405	.0165	.0002
5476	-.0015	.0036	.0080	-.0015
2660	-.0003	.1319	.0626	-.0003
2346	-.0042	.1018	-.0081	-.0039
2072	.0017	.2944	.0264	.0016
864	.0177	-.0308	.9972	.0134
609	.8478	.0003	-.0221	.5298
214	-.5299	-.0000	-.0020	.8480

Estimated Minima Ridge

Coded Radius	x^2	x^3	x^4	x^5	$Z(x)^*$
0.0	40.0	61.0	15.0	40.0	36182
0.5	39.7	65.7	17.5	46.1	34431
1.0	38.8	69.9	19.2	47.8	34310

Coded Radius	x^7	x^8	x^9	x^{10}	$Z(x)^*$
0.0	22.0	35.0	13.0	37.0	36182
0.5	22.3	38.9	13.1	37.6	34431
1.0	20.8	41.0	12.7	49.1	34310

Estimated Maxima Ridge

Coded Radius	x^2	x^3	x^4	x^5	$Z(x)^*$
0.0	40.0	61.0	15.0	40.0	36182
0.5	39.5	59.7	13.6	31.9	40332
1.0	39.0	60.7	12.9	24.6	47280

Coded Radius	x^7	x^8	x^9	x^{10}	$Z(x)^*$
0.0	22.0	35.0	13.0	37.0	36182
0.5	21.9	30.0	12.9	36.9	40332
1.0	21.9	22.9	12.8	36.9	47280

* - Regression estimate.

at terminals C and D or the tractors at terminal B. Furthermore, tests along the minimal ridge found multiple optimal solutions to a distance of .3 coded radius; at 1.0 the actual near-optimal solution of 35796 represents only a .79% increase over optimality. These results, combine with those of other known optimal solutions, provides the decision maker with a range of options in addition to the insight of the canonical analysis.

Regarding distributional analysis, Table 5.30 reports the results for selected optimal and near-optimal solutions from the response surface analysis and prior gradient searches. Unlike the previous problems, 4TERM exhibits a relatively stable and symmetric distribution, without either extremely high-cost (though relatively rare) scenarios, or detectable parameter or range changes within the region of optimality. Indeed, the near-optimal solution has the highest-cost z_k, suggesting that tradeoffs between lower maximum costs for slightly higher expected values does not occur in this problem, at least not along the minima ridge. Figure 5.4 gives the graphical presentation of 4TERM's distribution at optimality.

Finally, following the pattern in previous problems, the LH VRT turns in excellent results, while CVs produce a very mixed bag, with one case again showing an increase — 40% — over simple random sampling (Table 5.31).

The summary and follow-on analysis of these results can take on several forms and emphasis, depending upon the focus and interests of the decision-maker. One possibility would expand the minima ridge insight with a supplementary analysis of convex combinations of optimal solutions, or perhaps

TABLE 5.30
TOLERANCE LIMITS FOR 4TERM (RANDOM SEED = 58047800)

$x^1, \ldots, x^{10}$ ‡	Z(x)	Mdn. z_k	Population Min z_k	Population Max z_k	Tol. Limit Min z_k	Tol. Limit Max z_k	% Cvg.
144,40,61,15,40 93,22,35,13,37	35514	35767	22617	45962	22617	44322	.9844
137,47,61,15,40 108,24,35,8,25	35514	35767	22617	45962	22617	44322	.9844
151,32,60,15 78,19,35,19,49	35514	35767	22617	45962	22617	44322	.9844
124,39,70,19,48 76,21,41,13,49	35796	35767	22617	50578	22617	48938	.9922

‡ - $x^{11} \ldots x^{15}$ set to zero.

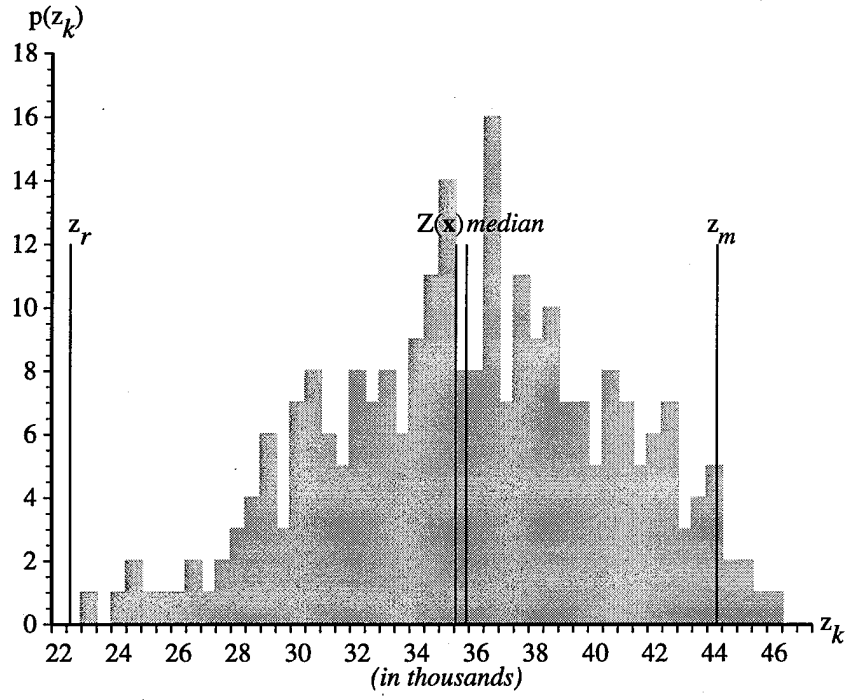


Figure 5.4. Comparison of Tolerance Limits to Population Distribution for 4TERM Centerpoint

TABLE 5.31
COMPARISON OF ESTIMATOR ACCURACY AND VARIANCE FOR
RS, CV, AND LH SAMPLING TECHNIQUES FOR 4TERM (I=50, N=10)

$x^1, \ldots, x^5$ $x^6, \ldots, x^{10}$ ‡	Z(x)	$\bar{Z}_{RS}$	$\bar{s}^2_{RS}$*	$\bar{Z}_{CV}$	$\bar{s}^2_{CV}$*	%†	$\bar{Z}_{LH}$	$\bar{s}^2_{LH}$*	%†
144,40,61,15,40 93,22,35,13,37	35514	35595	121	35507	76	.37	35522	20	.83
137,47,61,15,40 108,24,35,8,25	35514	35237	712	35246	536	.25	35515	6	.99
151,32,60,15 78,19,35,19,49	35514	35475	1170	35354	978	.16	35500	24	.98
137,40,63,16,44 90,22,38,13,37	35514	35196	680	35072	189	.72	35533	11	.98
124,39,70,19,48 76,21,41,13,49	35796	36165	1356	36271	1901	-.40	35768	99	.93

‡ - $x^{11} \ldots x^{15}$ set to zero. * - in thousands. † - % Variance reduction from RS ('-' implies increase).

finding additional response surface approximations using different optimal centerpoints. The examples below suggest two techniques for presenting the information.

1. *Ridge Charts.* This type of chart (Figures 5.5 and 5.6) captures the ridge results for the decision-maker by plotting the amount of change in the *uncoded* decision variables values per unit change in the *coded* deviation from the design centerpoint (e.g., optimal solution). For example, Figure 5.5 shows that if we wish to proceed half the distance away from the centerpoint in terms of the *coded* experimental design region, then terminal *D*'s tractor allocation would have to increase by 1, terminal *C*'s trailer allocation by 2, etc., to remain on the minima ridge. In essence, this chart simply displays in graphic form the data from Table 5.29.

2. *'Rule-of-Thumb'*. This type of table (Table 5.32) presents simple *notional* trend information for field use or operational guidance (i.e., given a choice, how do the drivers decide where to park the trucks overnight?).

Using these techniques, the multi-dimensional behavior of 4TERM can be summarized for the decision-maker as stated on the following page.

TABLE 5.32
NOTIONAL SUMMARY OF 4TERM SENSITIVITY*

	Terminal *B*		Terminal *C*		Terminal *D*	
	Trailer	Tractor	Trailer	Tractor	Trailer	Tractor
Near-Optimal	*Incr.*	*Incr.*	*Incr.*	—	*Incr.*	*Incr.*
Avoid	—	*Decr.*	*Decr.*	—	*Decr.*	—

* - Terminal *A* remains the same. Central node balances changes in Terminals *B*, *C*, *D*.

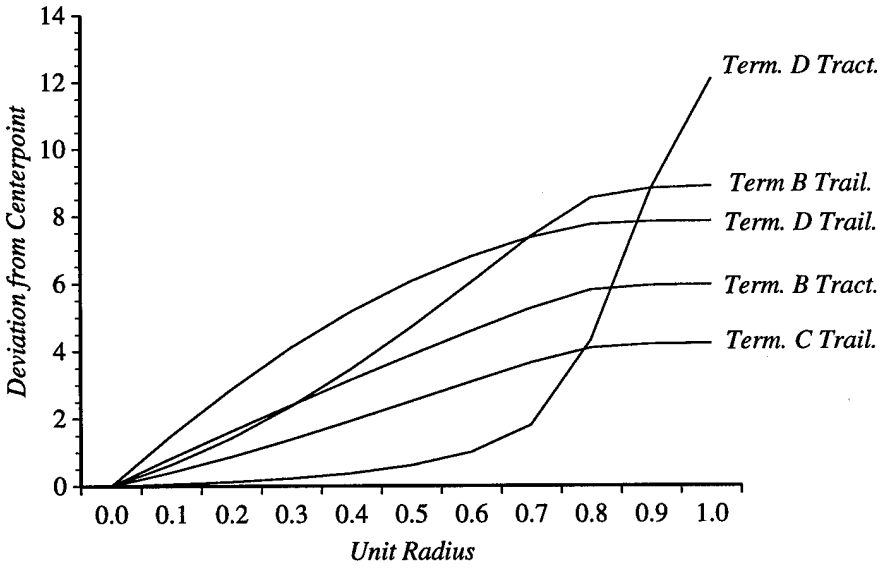


Figure 5.5. Minima Ridge Results for 4TERM

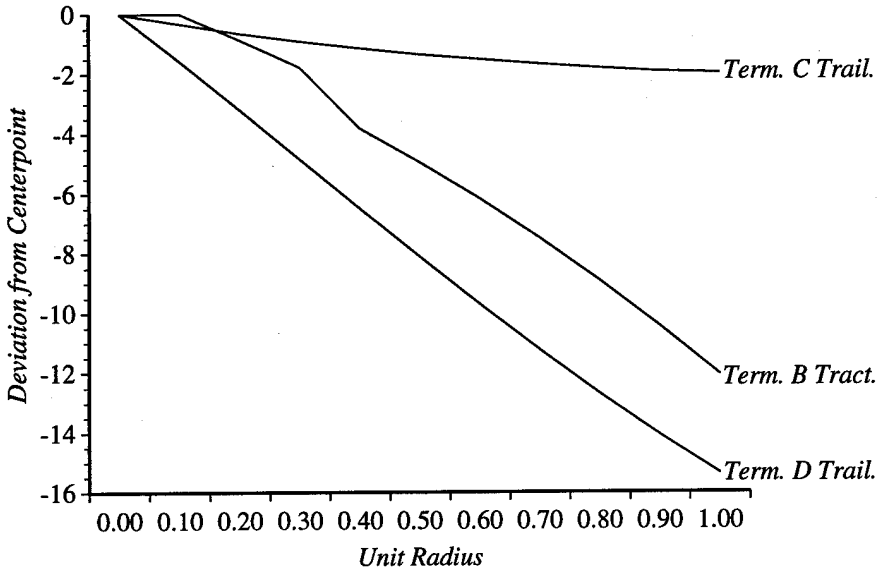


Figure 5.6. Maxima Ridge Results for 4TERM

4TERM Analysis Summary. *Figures 5.5 and 5.6 compare deviations from the proposed optimal solution that minimize and maximize, respectively, increases in expected cost, while Table 5.32 provides notional guidance. Optimal solutions should also minimize the highest-cost scenario.*

5.5 20TERM

5.5.1 20TERM Problem Description

20TERM, a straightforward extension of 4TERM, represents the largest stochastic problem this research investigates: 63 first-stage decision variables and 40 random right-side elements (again modeling pick-up and delivery demand for

the 20 outlying terminals). 20TERM's first-stage constraint structure $\mathbf{Ax} = \mathbf{b}$ follows the same pattern as 4TERM

$$\sum_{j=1}^{21} x^j = 600 \quad \textit{(Trailers)} \tag{5.4a}$$

$$\sum_{j=22}^{42} x^j = 400 \quad \textit{(Tractors)} \tag{5.4b}$$

$$\sum_{j=43}^{62} x^j \leq 10{,}000 \quad \textit{(Rental Tractor-Trailer)} \tag{5.4c}$$

where the cost of using $x^j, j = 43,\ldots, 62$ again is $c^j = 100$, while the existing fleet's expenses are zero; i.e., $c^j = 0, j = 1,\ldots, 42$. The $\mathbf{T}$ matrix converts the $\mathbf{x}$ variables to the recourse right-side in 4TERM's manner as well. As in the case of 4TERM, the right-side random variables each can take one of two values with equal probability; however, while 4TERM's eight variables allow for 256 total possible scenarios, 20TERM's 40 variables permit over $1.0995 \cdot 10^{12}$ distinct realizations of recourse demand. Finally, the size of 20TERM's recourse basis dimension increases to 128 from 4TERM's 28.

The resulting computational demands of 20TERM require modifying the previous response analysis strategy. The first obvious change recognizes that the true values for $Z(\mathbf{x})$ cannot be found (in any practical sense); therefore, both the search and experimental design must estimate $Z(\mathbf{x})$ using $\hat{Z}_s(\mathbf{x})$. However, exploratory samples of 20TERM show both a significant level of variance in $h(\mathbf{x},\omega,\mathbf{T})$ and VRT patterns similar to those seen already, thus ruling out $\hat{Z}_{RS}(\mathbf{x})$ and $\hat{Z}_{CV}(\mathbf{x})$. Furthermore, both preliminary tests on 20TERM and previous

problems indicate that the GEOMETRIC SIMPLEX and PARTAN ALGORITHMS are not up to the task of solving a problem of this size. Finally, tests on sample projected gradients most often found a unique basis for every sample realization of the random variables, thus rendering the OBS-RESET option ineffective. Consequently, this research resorts to finding a near-optimal solution with the PROJECTED GRADIENT ALGORITHM using the OSL option and the LH estimator $\hat{Z}_{LH}(\mathbf{x})$ with a stratification size of 200.

5.5.2 20TERM Optimization Results

Early research on applying the PROJECTED GRADIENT ALGORITHM to 20TERM found that, like the previous problems, the algorithm tends to find the region of optimality fairly quickly. However, unlike previous attempts, converging to a near-optimal solution requires far more computational time. Consequently, the following three modifications help adapt the PROJECTED GRADIENT ALGORITHM to the demands of 20TERM:

1. *Starting Solution.* Unlike previous applications of the PROJECTED GRADIENT ALGORITHM, where the initial solution represents an equal allocation of resources to each first-stage decision variable x^j, this problem uses the optimal solution ($\mathbf{x}_{ev}$) of the expected value approximation, with EV = MIN{$\mathbf{cx} + h(\mathbf{x},E[\omega],E[\mathbf{T}])$} as the starting point $\mathbf{x}_1$ (i.e., $\mathbf{x}_1 = \mathbf{x}_{ev}$). Recalling the estimator $\hat{Z}_{LH}(\mathbf{x}_{ev})$ equals the expected result of using the expected value solution $\mathbf{x}_{ev}$, $E[\hat{Z}_{LH}(\mathbf{x}_{ev})] \geq E[\hat{Z}_{LH}(\mathbf{x}^*)] \geq EV$. For 20TERM, EV= 239,272.9 while $\hat{Z}_{LH}(\mathbf{x}_{ev}) = 279,674.25$.

2. *Non-Constant Stratification Size.* The LH sampling size of the $\mathbf{x}_k$ vector that estimates the descent gradient is set to 200, whereas the stratification size of the responses fitted by the quadratic for calculating the stepsize drops to 50. This idea assumes that directional information coming from the gradient estimate of a *single* solution $\mathbf{x}_k$ should be more accurate than any single response estimate along the line projection.

3. *Constant Stepsize.* As the algorithm approaches near-optimality, the distance of the line segment defined by the directional gradient, incumbent solution $\mathbf{x}_k$, and lower bounds $\mathbf{x} \geq \mathbf{0}$ is such that its curvature becomes hard to detect. Therefore, at this point the PROJECTED GRADIENT ALGORITHM abandons equidistant sampling of the directional line segment in favor of small constant stepsizes over a set number of gradient iterations; in effect, resorting to Ermoliev's (1988) suggestion of following small, iterative search patterns.

Table 5.33 gives the results of the modified projected gradient search for a near optimal solution. After two iterations using quadratic estimates, the process shifts to using small constant stepsizes (.01-.04) over 50-100 iterations. The process terminates after run #13 indicates little additional progress being made after 100 iterations. It should be noted that while the projected gradients show continued descent possible, the 'leveling-off' trend suggests the true optimal solution will not be much less than 254,000; thus, the best solution found in run #13 should provide an adequate centerpoint ($\mathbf{x}_{CP}$) for fitting a response surface.

TABLE 5.33
PROJECTED GRADIENT RESULTS FOR 20TERM

Run #	Rand. Seed	Sclr. Type*	# Iter. (k)	CPU Time†	Start $\hat{Z}_{LH}(\mathbf{x})$	Best $\hat{Z}_{LH}(\mathbf{x})$‡
1	13414	Quadr. Est.	37	2473	279674.25	267810.03
2	885623	Quadr. Est.	14	450	267573.59	266883.09
3	64069848	Const. .01	50	1163	267508.91	266271.16
4	22691	Const. .03	50	1153	266719.00	264910.53
5	720760	Const. .04	100	2164	265302.31	264410.84
6	6273213	Const. .04	50	1105	265082.19	262471.28
7	931925	Const. .03	50	982	262697.75	261063.59
8	405692	Const. .02	75	1456	262228.81	259841.73
9	2555416	Const. .02	75	1452	259614.28	258648.66
10	598372248	Const. .03	75	1438	258730.72	257718.31
11	3956152	Const. .03	75	1409	258202.91	256692.53
12	4106067	Const. .03	75	1364	257749.39	255797.83
13	512255541	Const. .03	100	1836	256513.97	254945.70

* - $Q = 5$ for quadratic estimates. † - Units in seconds; total time 5 hrs. 7.5 mins.
‡ - Not necessarily k^{th} observation.

(Morton's (1994) results found the lower and upper bounds of $Z(\mathbf{x}^*)$ to be 249,747 and 256,497, respectively.) Refer to Table 5.34 in the following section for the values of $\mathbf{x}_{ev}$ and $\mathbf{x}_{CP}$.

5.5.3 20TERM Response Surface Analysis

Since 20TERM's number of first-stage decision variables renders their description in every table cumbersome, Table 5.34 consolidates the values of $\mathbf{x}_{ev}$, the best solution found by the PROJECTED GRADIENT ALGORITHM ($\mathbf{x}_{CP}$), and the factor ranges of the Plackett-Burman screening design. Table 5.34 also presents a good starting point for defining those factors and associated parameters for the experimental design phase of the analysis.

As discussed in Section 4.3, problems the size of 20TERM preclude using full factorial designs; indeed, even highly fractionated experimental designs for

TABLE 5.34
EXPECTED VALUE APPROXIMATION ($\mathbf{x}_{ev}$), EXPERIMENTAL DESIGN CENTERPOINT ($\mathbf{x}_{CP}$), AND SCREENING DESIGN (-,+) VALUES FOR 20TERM

Trailers

	x^1	x^2	x^3	x^4	x^5	x^6	x^7	x^8	x^9	x^{10}	x^{11}
$\mathbf{x}_{ev}$	136	20	50	8	0	0	21	0	0	22	29
$\mathbf{x}_{CP}$	78.5	25.5	55	13.5	0	0	26	0	0	27	34
-	—	20.5	50	8.5	—	—	21	—	—	22	29
+	—	30.5	60	18.5	—	—	31	—	—	32	39

	x^{12}	x^{13}	x^{14}	x^{15}	x^{16}	x^{17}	x^{18}	x^{19}	x^{20}	x^{21}
$\mathbf{x}_{ev}$	35	35	26	32	25	41	33	44	12	31
$\mathbf{x}_{CP}$	40.5	40	31	37	29.5	46	33	48	0	35.5
-	35.5	35	26	32	—	41	—	—	—	30.5
+	45.5	45	36	42	—	51	—	—	—	40.5

Tractors

	x^{22}	x^{23}	x^{24}	x^{25}	x^{26}	x^{27}	x^{28}	x^{29}	x^{30}	x^{31}	x^{32}
$\mathbf{x}_{ev}$	167.5	10	25	4	0	0	10.5	0	0	11	18.5
$\mathbf{x}_{CP}$	129.5	13.5	28.5	6.5	0	0	13.5	0	0	14	18
-	—	—	—	—	—	—	—	—	—	—	—
+	—	—	—	—	—	—	—	—	—	—	—

	x^{33}	x^{34}	x^{35}	x^{36}	x^{37}	x^{38}	x^{39}	x^{40}	x^{41}	x^{42}
$\mathbf{x}_{ev}$	17.5	17.5	13	16	12.5	23	16.5	22	0	15.5
$\mathbf{x}_{CP}$	20	20	15.5	19	17	24.5	17	25	0	18.5
-	—	—	—	—	12	—	—	—	—	—
+	—	—	—	—	22	—	—	—	—	—

Tractor-Trailer Rental

	x^{43}	x^{44}	x^{45}	x^{46}	x^{47}	x^{48}	x^{49}	x^{50}	x^{51}	x^{52}	x^{53}
$\mathbf{x}_{ev}$	312	0	0	0	30	38	0	14	12	0	0
$\mathbf{x}_{CP}$	382.5	0	0	0	35.5	42.5	0	19	17.5	0	0
-	377.5	—	—	—	30.5	37.5	—	14	12.5	—	—
+	387.5	—	—	—	40.5	47.5	—	24	22.5	—	—

	x^{54}	x^{55}	x^{56}	x^{57}	x^{58}	x^{59}	x^{60}	x^{61}	x^{62}	x^{63}	x^{64}
$\mathbf{x}_{ev}$	0	0	0	0	0	0	0	0	12	0	9582
$\mathbf{x}_{CP}$	0	0	0	0	0	0	0	0	28.5	0	9475
-	—	—	—	—	—	—	—	—	23.5	—	—
+	—	—	—	—	—	—	—	—	33.5	—	—

the number of factors present in 20TERM are prohibitive. Consequently, reducing the number of factors based on subjective interests of the decision-maker, *a priori* knowledge or insight into the problem, or preliminary screening

becomes necessary. This research employs the last technique by selecting the nineteen x^j variables that differ the most between $\mathbf{x}_{ev}$ and $\mathbf{x}_{CP}$ as candidates for a Plackett-Burman screening design suggested by Montgomery (1984) and Plackett and Burman (1946). This criteria implies an interest in those factors that we wish to *avoid changing*. By contrast, screening for those factors that *change the least* during the search assumes a greater interest in finding alternative near-optimal solutions. In this instance, although 36 variables change value from starting to near-optimal solutions, the most significant differ by 5 or more. As in the case of 4TERM, variables x^1 and x^{22} provide the necessary degree of freedom for the equality constraints; hence, they're ignored. Table 5.35 shows the factor settings and estimated responses, while Table 5.36 provides the regression analysis.

The coded parameter estimates provide the basis for selecting with factors to include in a CCD design. Following the selection criteria of using those variables that most influence the response, the following 11 variables represent a descending order of first-order significance: x^4, x^{14}, x^7, x^{21}, x^{37}, x^{13}, x^{51}, x^{10}, x^{62}, x^{12}, and x^2. Although x^{11} could arguably be included (its parameter estimate is only 10 less than x^2), CCD design limitations restrict the number of factors to 11 in order to keep the fractional portion to 128 runs and still retain a resolution V level.

The final fractional portion of the CCD design follows a design generator suggested by Lorenzen and Anderson (1993) where a full-factorial design for seven variables — x^4, x^7, x^{14}, x^{21}, x^{37}, x^{51}, x^{62} — provides the structure for the $2^{11-4} = 128$ runs. The remaining factors' coded values are calculated as follows:

TABLE 5.35
PLACKETT-BURMAN SCREENING DESIGN FOR 20TERM

*Selected x_j Variables From Table 5.34**																		$\hat{Z}_{LH}(\mathbf{x})$
+	+	-	-	+	+	+	+	-	+	-	+	-	-	-	-	+	+	-	272044
-	+	+	-	-	+	+	+	+	-	+	-	+	-	-	-	-	+	+	268636
+	-	+	+	-	-	+	+	+	+	-	+	-	+	-	-	-	-	+	269038
+	+	-	+	+	-	-	+	+	+	+	-	+	-	+	-	-	-	-	271066
-	+	+	-	+	+	-	-	+	+	+	+	-	+	-	+	-	-	-	270387
-	-	+	+	-	+	+	-	-	+	+	+	+	-	+	-	+	-	-	269650
-	-	-	+	+	-	+	+	-	-	+	+	+	+	-	+	-	+	-	271463
-	-	-	-	+	+	-	+	+	-	-	+	+	+	+	-	+	-	+	271990
+	-	-	-	-	+	+	-	+	+	-	-	+	+	+	+	-	+	-	272466
-	+	-	-	-	-	+	+	-	+	+	-	-	+	+	+	+	-	+	274566
+	-	+	-	-	-	-	+	+	-	+	+	-	-	+	+	+	+	-	269603
-	+	-	+	-	-	-	-	+	+	-	+	+	-	-	+	+	+	+	271065
+	-	+	-	+	-	-	-	-	+	+	-	+	+	-	-	+	+	+	271042
+	+	-	+	-	+	-	-	-	-	+	+	-	+	+	-	-	+	+	272404
+	+	+	-	+	-	+	-	-	-	-	+	+	-	+	+	-	-	+	269194
+	+	+	+	-	+	-	+	-	-	-	-	+	+	-	+	+	-	-	270591
-	+	+	+	+	-	+	-	+	-	-	-	-	+	+	-	+	+	-	270286
-	-	+	+	+	+	-	+	-	+	-	-	-	-	+	+	-	+	+	269377
+	-	-	+	+	+	+	-	+	-	+	-	-	-	-	+	+	-	+	271671
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	284791

* - Columns correspond to variables in Table 5.34 (from left to right in ascending order of j) for those variables with values for '-' and '+'. For instance, the second column represents x^3 where '-' equals 50 and '+' equals 60. Random Seed for this design is 630011823.

$$x^2 = x^{21} \cdot x^{37} \cdot x^{51} \cdot x^{62} \tag{5.5a}$$

$$x^{10} = x^7 \cdot x^{14} \cdot x^{51} \cdot x^{62} \tag{5.5b}$$

$$x^{12} = x^4 \cdot x^{14} \cdot x^{37} \cdot x^{62} \tag{5.5c}$$

$$x^{13} = x^4 \cdot x^7 \cdot x^{14} \cdot x^{21} \cdot x^{37} \cdot x^{51} \cdot x^{62}. \tag{5.5d}$$

The next step of determining the axial values and number of centerpoints requires special consideration due to the variability of the response estimator $\hat{Z}_{LH}(\mathbf{x})$. As related by Lorenzen and Anderson (1993), where F represents the size of the fractional portion ($F = 2^{K-P}$), $2K$ the number of axial points, and n_K

TABLE 5.36
REGRESSION RESULTS FOR PLACKETT-BURMAN DESIGN IN TABLE 5.35

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	R Square
Model	19	223527868	11764625	1.0
Error	0			
Total	19	223527868		

Selected Parameter Estimates

Variable	Coded Par. Est.	Uncod. Par. Est.
Intercept	271566	359928
x^2	-654.63	-130.93
x^3	-542.70	-108.54
x^4	-1786.22	-357.24
x^7	-905.39	-181.08
x^{10}	-714.49	-142.90
x^{11}	-644.85	-128.99
x^{12}	-665.11	-133.02
x^{13}	-729.11	-145.82
x^{14}	-945.78	-189.16
x^{15}	-496.35	-99.27
x^{17}	-517.94	-103.53
x^{21}	-882.77	-176.55
x^{37}	-850.18	-170.04
x^{43}	-143.27	-28.65
x^{47}	-506.37	-101.27
x^{48}	-528.24	-105.65
x^{50}	-315.83	-63.17
x^{51}	-727.77	-145.55
x^{62}	-668.19	-133.64

the number of centerpoint replications, a CCD design becomes orthogonal by selecting the axial coded multiplier α_O using the relations

$$Q = (\sqrt{F + 2K + n_K} - \sqrt{F})^2 \tag{5.6a}$$

$$\alpha_O = \sqrt[4]{QF/4}, \tag{5.6b}$$

whereas the axial multiplier α_R for a rotatable design is

$$\alpha_R = \sqrt[4]{F}. \tag{5.7}$$

Since F and K are fixed, it follows immediately from (5.6) and (5.7) that n_K should be selected such that Q equals 4 for a design to have both properties. Applying this result to the proposed design for 20TERM gives the equivalent expression of finding the integer n_K such that

$$n_K \approx (2 + \sqrt{128})^2 - 150, \tag{5.8}$$

or $n_K = 27$. Plugging n_K back into (5.6) gives the final design's orthogonal axial multiplier α_O as 3.3555, compared to the rotatable multiplier $\alpha_R = 3.3636$.

The appendix presents the experimental design results using the three primary estimators $\hat{Z}_{RS}(\mathbf{x})$, $\hat{Z}_{CV}(\mathbf{x})$, and $\hat{Z}_{LH}(\mathbf{x})$. Table 5.37 on the following pages gives the regression results for the most significant parameters, with the linear and quadratic terms dominating the polynomial approximation. Table 5.38 provides the A canonical analysis results.

Since the Plackett-Burman screening design focuses on those x^js that influence the estimated response $\hat{Z}_{LH}(\mathbf{x})$ *the most*, the resulting canonical analysis provides an excellent estimate of the maxima ridge. By contrast, while Table 5.38 does provide a minima ridge assessment, a better estimate of the direction of minimum sensitivity can be found by re-accomplishing the preceding steps with a screening design composed of those x^js that affect the estimated response $\hat{Z}_{LH}(\mathbf{x})$ *the least*. Therefore, this summary will forgo a minima ridge recommendation, and concentrate instead on characterizing the most influential components of $\mathbf{x}$.

TABLE 5.37

REGRESSION RESULTS FOR EXPERIMENTAL DESIGN IN APPENDIX FOR 20TERM*

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F-Ratio	R Square
Model	73	1288589185	17651907	127	.8903
Error	26	3612523	138943		
Total	99	1292201709			

Selected Parameter Estimates

Variable	Uncod. Par. Est.	Std. Error	Cod. Par. Est.
Intercept	685077.00	33499.00	255528.00
x^2	-2917.43	540.06	-43178.00
x^4	-4077.62	1012.075	-4181.95
x^7	-2737.28	540.84	-2048.05
x^{10}	-3199.12	542.45	-3526.97
x^{12}	-3140.76	411.21	-4720.38
x^{13}	-2970.91	410.70	-3568.97
x^{14}	-2093.84	549.42	-1387.97
x^{21}	-3235.77	406.38	-5210.70
x^{37}	-2060.16	810.09	-2774.82
x^{51}	-2275.19	811.86	-2099.77
x^{62}	-1157.21	544.96	-1610.20
$x^2 \cdot x^2$	32.41	4.03	20531.00
$x^4 \cdot x^4$	65.56	14.18	11814.00
$x^7 \cdot x^7$	25.08	4.03	15892.00
$x^{10} \cdot x^{10}$	33.60	4.03	21289.00
$x^{12} \cdot x^{12}$	23.79	2.27	26791.00
$x^{12} \cdot x^{21}$	6.40	3.19	7213.03
$x^{13} \cdot x^{13}$	20.62	2.29	23221.00
$x^{13} \cdot x^{21}$	7.07	3.19	7959.89
$x^{14} \cdot x^{14}$	30.64	4.03	19410.00
$x^{21} \cdot x^{21}$	25.16	2.27	28337.00
$x^{37} \cdot x^{37}$	41.43	9.07	11665.00
$x^{51} \cdot x^{51}$	38.39	9.07	10810.00
$x^{62} \cdot x^{62}$	17.39	4.03	11016.00

* - Based on $\hat{Z}_{LH}(\mathbf{x})$ estimator.

TABLE 5.38
A CANONICAL ANALYSIS OF 20TERM

			Eigenvectors			
Eigenvalues	x^2	x^4	x^7	x^{10}	x^{12}	x^{13}
38670	.2619	.1484	.2363	.3093	.4796	.4061
23945	.0439	.0037	.0255	-.0581	.7836	-.0340
21508	.2778	.0783	.1317	.2816	-.3619	.6327
19842	.4229	.0638	.0228	-.4375	-.0031	-.1586
19503	.1807	.0712	.0365	.7517	-.0358	-.5735
18848	-.7904	.0873	.0169	.1558	.0804	.1588
14798	-.1004	.0196	.9207	-.1198	-.1100	-.1939
12528	.0201	.5770	-.0066	-.1349	-.0025	-.0928
11136	.0214	-.3849	.0413	.0061	.0253	.0125
10399	.0717	-.4101	-.1688	.0888	-.0011	-.0485
9597	-.0065	.5526	-.2146	.0217	-.0672	-.0557

Eigenvalues	x^{14}	x^{21}	x^{37}	x^{51}	x^{62}
38670	.0061	.5929	.0367	.0864	.0119
23945	.0282	-.6124	-.0464	.0306	-.0002
21508	.2054	-.4931	.0533	.0121	.0198
19842	.7636	.1177	-.0370	.0367	-.0155
19503	.2207	-.0667	.0123	-.1110	.0091
18848	.5472	.0456	.0818	.0549	.0177
14798	-.0846	-.0763	-.0309	.2525	.0035
12528	-.1023	-.0485	.7625	-.1071	.1845
11136	.0467	.0250	.0706	-.0692	.9145
10399	.0252	-.0080	.4598	.7444	-.1468
9597	-.0728	-.0326	-.4320	.5835	.3270

			Estimated Minima Ridge				
Coded Radius	x^2	x^4	x^7	x^{10}	x^{12}	x^{13}	$Z(\mathbf{x})^*$
0.0	25.5	13.5	26.0	27.0	40.5	40.0	255528
1.0	28.9	21.6	21.4	28.6	41.2	38.6	261453

Coded Radius	x^{14}	x^{21}	x^{37}	x^{51}	x^{62}	$Z(\mathbf{x})^*$
0.0	31.0	35.5	17.0	17.5	28.5	255528
1.0	30.6	36.4	18.9	28.2	38.5	261453

			Estimated Maxima Ridge				
Coded Radius	x^2	x^4	x^7	x^{10}	x^{12}	x^{13}	$Z(\mathbf{x})^*$
0.0	25.5	13.5	26.0	27.0	40.5	40.0	255528
1.0	18.1	10.9	20.2	19.0	24.6	26.7	304841

Coded Radius	x^{14}	x^{21}	x^{37}	x^{51}	x^{62}	$Z(\mathbf{x})^*$
0.0	31.0	35.5	17.0	17.5	28.5	255528
1.0	30.1	16.4	15.7	15.8	27.6	304841

* - Regression estimate.

In that context, Table 5.38 shows that reducing x^2, x^7, x^{10}, x^{12}, x^{13}, and x^{21} from their $\mathbf{x}_{CP}$ values considerably increases the estimated response $\hat{Z}_{LH}(\mathbf{x})$. Examining the eigenvectors reveals these factors as prominent components in the rotated axes with the highest eigenvalues, although in this design every axis exhibits significant curvature (again reflecting the screening design's choices). Figure 5.7 expresses this phenomenon in graphical terms for easier understanding.

Unlike the previous problems, 20TERM does not afford the true population parameters for comparison to sample-based estimates. Furthermore, the current design's emphasis on influential factors suggests little likelihood of finding lower maximum values of z_k at any location other than $\mathbf{x}_{CP}$. Therefore, this analysis presents the tolerance limits for $\mathbf{x}_{ev}$ and $\mathbf{x}_{CP}$ in Table 5.39 and a histogram of 400 random samples of z_k at $\mathbf{x}_{CP}$ in Figure 5.8. Both the tolerance limit results and sample distribution suggest 20TERM follows a near-symmetrical distribution similar to 4TERM. These results suggest the following analysis summary.

20TERM Analysis Summary. *Suggest using $\mathbf{x}_{CP}$ as defined in Table 5.34. Avoid reducing the current values of the decision variables as shown in Figure 5.7; however, increases in these figures can occur with small gains in $Z(\mathbf{x})$. Tolerance limits suggest near-symmetrical distribution with upper limit approximately \$50k over expected value. Minimal ridge estimation requires further analysis.*

TABLE 5.39
TOLERANCE LIMITS FOR 20TERM (RANDOM SEED = 3623643)

$\mathbf{x}_k$	$\hat{Z}_{RS}(\mathbf{x}_k)$	*Tolerance Limit* Lower Limit z_r	Upper Limit z_m
$\mathbf{x}_{ev}$	283596	239281	339276
$\mathbf{x}_{CP}$	258557	230766	307655

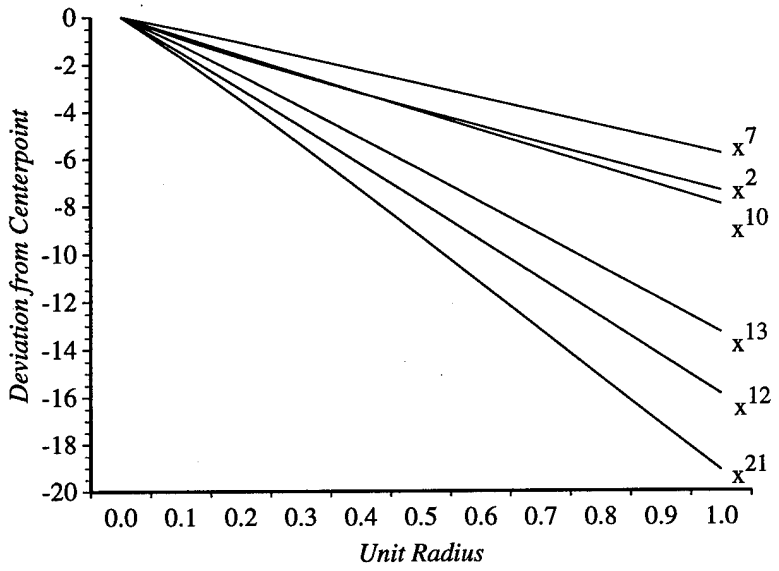


Figure 5.7. Maxima Ridge Results for 20TERM

Finally, although this research uses LH sampling only for the previous analysis of 20TERM, applying RS and CV sampling to (1) the centerpoint portion of the experimental design and (2) the entire final design in the appendix provides ways to measure of the amount of variance reduction using LH sampling. First,

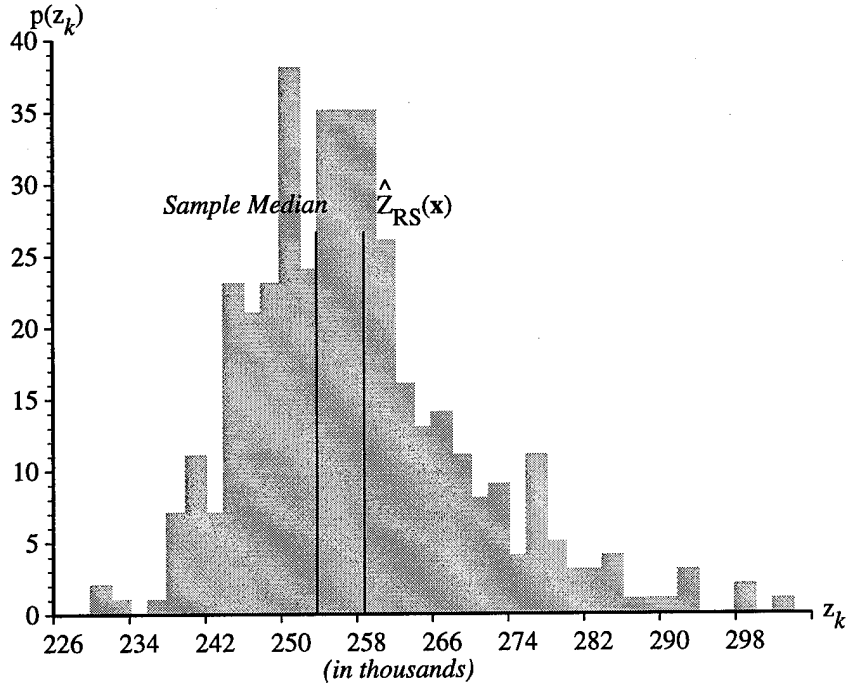


Figure 5.8. Sample Distribution of z_k for Centerpoint ($\mathbf{x}_k$)

using the 27 estimators of the centerpoint in the same fashion as the previous problems use 10 estimators of a given $Z(\mathbf{x}_k)$ gives the results shown in Table 5.40. Since the sample variance estimators in Table 5.40 are equivalent to the mean square pure error under linear regression, a natural extension of such an analysis would compare the regression results using different sampling techniques for the *entire design*. Table 5.41 provides such a comparison (note that both the *lack of fit* and *pure error* drops when using VRT — especially LHs).

TABLE 5.40
COMPARISON OF ESTIMATOR ACCURACY AND VARIANCE FOR
RS, CV, AND LH SAMPLING TECHNIQUES FOR 20TERM (I=50, N=27)

$x^2, x^4, x^7, x^{10}, x^{12}, x^{13}$ $x^{14}, x^{21}, x^{37}, x^{51}, x^{62}$											$\hat{Z}_{RS}(x_0)$	$\hat{Z}_{CV}(x_0)$	$\hat{Z}_{LH}(x_0)$
0	0	0	0	0	0	0	0	0	0	0	256393.22	255784.72	256009.81
0	0	0	0	0	0	0	0	0	0	0	255674.83	255718.84	254763.36
0	0	0	0	0	0	0	0	0	0	0	253818.95	253197.25	255341.89
0	0	0	0	0	0	0	0	0	0	0	253555.20	253433.41	254785.86
0	0	0	0	0	0	0	0	0	0	0	252856.48	252810.88	255575.00
0	0	0	0	0	0	0	0	0	0	0	255490.61	255577.50	255235.52
0	0	0	0	0	0	0	0	0	0	0	252902.50	254253.78	255041.83
0	0	0	0	0	0	0	0	0	0	0	253241.92	254443.30	255877.17
0	0	0	0	0	0	0	0	0	0	0	256394.80	256144.66	255931.16
0	0	0	0	0	0	0	0	0	0	0	254077.73	253923.22	255685.38
0	0	0	0	0	0	0	0	0	0	0	254993.14	254751.05	255091.02
0	0	0	0	0	0	0	0	0	0	0	257465.98	257531.95	255311.55
0	0	0	0	0	0	0	0	0	0	0	254823.16	255137.92	255446.63
0	0	0	0	0	0	0	0	0	0	0	255125.48	254839.91	255863.88
0	0	0	0	0	0	0	0	0	0	0	255648.98	255481.08	255543.17
0	0	0	0	0	0	0	0	0	0	0	258436.55	257506.67	254989.86
0	0	0	0	0	0	0	0	0	0	0	257079.67	255383.59	255698.77
0	0	0	0	0	0	0	0	0	0	0	254280.70	254184.72	254915.28
0	0	0	0	0	0	0	0	0	0	0	252949.27	252760.42	255062.73
0	0	0	0	0	0	0	0	0	0	0	260193.55	260242.17	255693.64
0	0	0	0	0	0	0	0	0	0	0	256091.58	256139.03	255246.78
0	0	0	0	0	0	0	0	0	0	0	257106.16	254932.30	255796.36
0	0	0	0	0	0	0	0	0	0	0	254390.89	254333.63	255222.17
0	0	0	0	0	0	0	0	0	0	0	255581.75	255756.41	255611.36
0	0	0	0	0	0	0	0	0	0	0	255913.77	256291.66	255996.63
0	0	0	0	0	0	0	0	0	0	0	257371.08	258470.84	255658.77
0	0	0	0	0	0	0	0	0	0	0	254371.97	254671.95	255388.92

	$\hat{Z}_{RS}(x_0)$	$\hat{Z}_{CV}(x_0)$	$\hat{Z}_{LH}(x_0)$
Estimated Mean $\bar{Z}(x_0)$	255415.92	255322.33	255436.46
Sample Variance $\bar{s}^2(x_0)$	3226848	2883646	138943
% Variance Reduction	—	.1064	.9569

TABLE 5.41
REGRESSION RESULTS FOR EXPERIMENTAL DESIGN IN APPENDIX FOR 20TERM

Random Sampling

Residual	DF	Sum of Squares	Mean Square	F-Ratio	R Square
Lack of Fit	73	1441633317	19748402	6.12	.8757
Pure Error	26	83898043	3226848		
Total Error	99	1525531360	15409408		

Control Variates

Residual	DF	Sum of Squares	Mean Square	F-Ratio	R Square
Lack of Fit	73	1426942064	19547152	6.78	.8787
Pure Error	26	74974798	2883646		
Total Error	99	1501916862	15170877		

Latin Hypercube

Residual	DF	Sum of Squares	Mean Square	F-Ratio	R Square
Lack of Fit	73	1288589185	17651907	127.00	.8903
Pure Error	26	3612523	138943		
Total Error	99	1292201709	13052543		

Chapter 6

Conclusions

6.1 INTRODUCTION

This chapter closes this dissertation by reviewing its research efforts in the following format:

1. *Results and Contributions.* Surveys the computational requirements and empirical results of the proposed techniques for finding $\mathbf{x}^*$, deriving a polynomial approximation of $Z(\mathbf{x})$ as a function of $\mathbf{x}$, and integrating the underlying distributional characteristics in the decision process. Also examines this dissertation's contributions — and their significance — to the topic of two-stage stochastic linear programming with recourse.

2. *Recommendations for Future Research.* Suggests areas of future research based on the discoveries of this study.

3. *Conclusions.* Summarizes this dissertation's accomplishments.

Each section follows the same organizational format as the rest of this dissertation regarding the topics listed under *Optimization Methods* and *Statistical Analysis*. Figure 6.1 on the following page provides a summary chart of this chapter for quick reference.

	Search Technique	*Optimization Algorithm*	*Variance Reduction*	*Experimental Design*	*Response Surface Analysis*	*Distribution Analysis*
Results	**Simplex** Premature convergence **Proj. Gradient** Excellent **PARTAN** Slow, but convergent	**OSL-COMPL./ ODV** Excellent; Available only on small prob. **OSL-RESET** Good; Works on all but largest prob.	**Control Variates** Inconsistent **Latin HyperCube** Excellent variance reduction	**Preliminary / Central Composite** Demonstrated viability of applying exper. design tech. on recourse prob.	**Canonical / Ridge Analysis** Found good to excellent fits for all prob.; Pos. Def.; Excellent insight using Min/Max ridge	**Tolerance Limits** One case of lower upper limits for higher $Z(\mathbf{x})$; Most distrb. heavily skewed
Further Research	**Proj. Gradient** Adaptive algorithm	**OSL-RESET** Expand beyond specific $\mathbf{x}_k$ Dynamic mgt.	**Combined VRT** Integrate Latin hypercube with other VRT	**Other Designs** Minimum Bias; Box-Behnken; Mixture; Factorial Sampling	**Resp. Analysis** Additional responses;	**Non-Parametric** Quantile; Skewness

Figure 6.1. Summary of Dissertation Results

6.2 RESULTS AND CONTRIBUTIONS

6.2.1 Optimization Methods

This dissertation investigated several techniques in two basic categories for finding $\mathbf{x}^*$ as the starting point for conducting response surface analyses — *search methods* (projected gradient, geometric simplex, and PARTAN), and *optimization algorithms* (optimal bases and dual vector sets). Regarding search methods, the PROJECTED GRADIENT ALGORITHM clearly outperforms the other two algorithms for all problems in terms of computational duration and convergence; indeed, neither the GEOMETRIC SIMPLEX or PARTAN ALGORITHMS can handle larger problems (20TERM) in any reasonable amount of time. The PROJECTED GRADIENT ALGORITHM's clearest advantage lies in its ability to find an accurate directional descent vector; neither the straightforward GEOMETRIC SIMPLEX or PARTAN's parallel tangent property gave better directional guidance for the computational time either saved or expended, respectively. Furthermore, using the quadratic fit of the response along the directional descent for estimating the stepsize gives the PROJECTED GRADIENT ALGORITHM the capability to find the region of optimality fairly quickly *even for the largest problems*; only in the region of optimality for those cases does it become more tractable to resort to predetermined stepsizes.

As for the OBS-COMPLETE and ODV techniques, such methods provide clear computational advantages for smaller recourse problems. The OBS-RESET option also proves advantageous over repetitive OSL calls for medium-sized

problems, but not the order of magnitude seen with the smaller ones. Ultimately, though, this technique is problem-dependent, and for very large problems may not be a viable option.

6.2.2 Statistical Analysis

The results from the response surface approximation of $Z(\mathbf{x})$ establish the viability and usefulness of this form of analysis for two-stage stochastic linear programming with recourse. Specifically, the results and contributions of this research in this category are summarized below:

1. *Variance Reduction.* This dissertation establishes that Latin hypercube sampling *guarantees* a reduction in the variance of the sample estimator of $Z(\mathbf{x})$ over random sampling for two-stage capacity expansion problems, and empirically confirms such reductions as both large and consistent for the set of test problems. *Most importantly, this variance reduction technique can be applied to any algorithm or analytical technique that employs statistical estimation of the objective function for two-stage stochastic linear programming problems with recourse.* By contrast, using the random elements of $\mathbf{T}$ and ω as control variates generally does not reduce the variance nearly as well as the Latin hypercube technique; indeed, cases exist where such controls *increase* the variance of the sample estimator. Furthermore, as a practical matter using Latin hypercube sampling demands very few additional computations, and — unlike

control variates — requires no knowledge or guesses on correlation to the response $h(\mathbf{x},\omega,\mathbf{T})$.

2. *Experimental Design.* This study demonstrates that experimental design techniques — such as preliminary factor screening, fractional design, and orthogonal, rotatable central composite designs — can be successfully applied to this class of problems.

3. *Response Surface Analysis.* This research demonstrates the feasibility of fitting a second-order polynomial to $Z(\mathbf{x})$ in the region of optimality. Although sometimes requiring factor adjustments in range or centerpoints, all problems in the test set can be fit with a positive definite quadratic form and R^2 factor near .9 or better. *Most importantly, the canonical analysis of these approximations empirically confirms the existence of optimal or near-optimal regions, and provides a method of sensitivity analysis not available until now.*

4. *Tolerance Limits.* Finally, this dissertation applies the non-parametric technique of tolerance limits to characterize the underlying distribution, and to incorporate such results in the decision-making process. Although problem-dependent, such analysis found cases in the current problem set where either pathologically skewed distributions or reduced tolerance ranges for near-optimal solutions suggest expanding the decision criteria beyond MIN $Z(\mathbf{x})$.

6.3 RECOMMENDATIONS FOR FUTURE RESEARCH

6.3.1 Optimization Methods

The disappointing results of the GEOMETRIC SIMPLEX (and to a lesser extent PARTAN) ALGORITHMS offer limited possibilities for further research in these areas; both their performance and inherent liabilities in the stochastic recourse environment suggest little likelihood of improvement. By contrast, the PROJECTED GRADIENT ALGORITHM proves itself to be a viable method for finding the optimal or near-optimal solution for even the largest problems, and perhaps can be improved upon in the following areas:

1. *Stepsize Estimation.* This dissertation shows that using a quadratic estimate of $Z(\mathbf{x})$ as a function of the projected gradient multiplier very quickly finds the region of near-optimality; however, the method does not always work, especially as the search nears optimality. Further investigations into different approximation methods may increase its accuracy.

2. *Adaptive Search Techniques.* As implemented, the PROJECTED GRADIENT ALGORITHM uses constant parameters for the number of points searched along the line segment, their sampling size, and the length of the line segment itself. However, this 'one-size-fits-all' approach clearly does not work as efficiently in the region of optimality, again particularly for larger problems. Consequently, a more dynamic approach whereby the algorithm adjusts the search process in order to gain more precise

information in the immediate area of the incumbent solution should provide better results.

Regarding optimal basis sets, their use clearly provides computational advantages whenever the specific problem makes them available. Unfortunately, this research strongly suggests that the OBS-COMPLETE technique will not be practical for larger problems. However, this study does not fully explore cases where the OBS-RESET method might prove practical for even larger problems. These extensions include:

1. *Expanded Reset Option.* The current algorithm resets the optimal basis set for each feasible $\mathbf{x}_k$, regardless of the specifics of the algorithm in use; however, another feasible $\mathbf{x}_{k+1}$ 'nearby' may share a significant number of optimal bases. For instance, the proposed reduced line segment in a revised PROJECTED GRADIENT ALGORITHM may require a reasonably small number of optimal bases along its entire length; in such a case, the short segments of a problem like 20TERM in its region of optimality could be estimated more quickly. This same phenomenon might occur in experimental design settings as well. Obviously, repetitive sampling of a single point (such as the centerpoint) would benefit; however, with the reduced number of factors common to fractional designs, a single optimal basis set might still be practical. Furthermore, such a case would allow replications *at all design points* (not just the centerpoints), providing even better estimates of experimental error.

2. *Dynamic Reset.* Another suggestion would be to 'sort-and-trim' the optimal basis set as an integral part of the iterations of any particular algorithm. Based on frequency of optimality, as less-used bases progressively move to the bottom of the list they would be replaced by newer, more frequently used optimal bases in an on-going process.

6.3.2 Statistical Analysis

Just as this dissertation's principal contributions lie in applying statistical analysis techniques — variance reduction, response surface analysis, and non-parametric statistics — to the recourse problem, so do the most interesting avenues for further research. Specifically, these include the following suggestions:

1. *Variance Reduction.* While the Latin hypercube technique substantially lowers the variance of the estimators of $Z(\mathbf{x})$, even further variance reduction may be possible through its combined use with other VRTs. One particularly promising prospect involves using a single control variate proposed by Morton (1995b).

2. *Response Surface Analysis.* This research employs only basic experimental designs and response surface techniques to describe $Z(\mathbf{x})$. Additional areas of research include using minimum bias designs, experimental design structures other than central composite designs, and preliminary factor sampling (Morris 1991). Furthermore, additional

polynomial approximations to responses other than $Z(\mathbf{x})$ may prove useful as well.

3. *Distributional Analysis.* Additional non-parametric analysis, such as quantile, median, and skewness estimates, would further characterize the underlying distribution of $h(\mathbf{x},\omega,\mathbf{T})$. As with tolerance limits, such information would provide additional insight to the decision-maker.

6.4 CONCLUSIONS

Since its introduction 40 years ago, researchers have devoted considerable theoretical and empirical research into understanding and solving two-stage stochastic linear programming with recourse. During this same period simulation — including the related fields of experimental design, variance reduction, and response surface methodology — developed into a powerful method of analysis for problems inherently stochastic in nature. This dissertation represents a formal *synthesis* of these two fields — an investigation in how to apply the methods of one to get answers and insight about the other. In so doing it brings a new philosophy to solving an old problem while opening additional avenues of research. It accomplishes this from a *tactical* point-of-view by providing new techniques for efficiently and accurately solving the classic optimization problem. Most importantly, from a *strategic* perspective this research introduces the equally important topics of sensitivity and distributional analysis by demonstrating their viability with respect to this class of stochastic linear programming problems.

Appendix

TABLE A.1
EXPERIMENTAL DESIGN FOR 20TERM†

$x^2, x^4, x^7, x^{10}, x^{12}, x^{13}$ $x^{14}, x^{21}, x^{37}, x^{51}, x^{62}$											$\hat{Z}_{RS}(\mathbf{x}_k)$	$\hat{Z}_{CV}(\mathbf{x}_k)$	$\hat{Z}_{LH}(\mathbf{x}_k)$
1	-1	-1	1	1	-1	-1	-1	-1	-1	-1	277197.81	277058.50	276014.53
1	1	-1	1	-1	1	-1	-1	-1	-1	-1	275254.03	275498.84	272648.09
1	-1	1	-1	1	1	-1	-1	-1	-1	-1	272360.13	272067.63	273361.38
1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	279219.50	278849.44	277337.97
1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	280607.00	280422.78	279328.13
1	1	-1	-1	1	-1	1	-1	-1	-1	-1	276574.34	275890.34	275684.44
1	-1	1	1	-1	-1	1	-1	-1	-1	-1	279598.56	278392.69	277554.34
1	1	1	1	1	1	1	-1	-1	-1	-1	274301.03	275076.59	272743.44
-1	-1	-1	1	1	1	-1	1	-1	-1	-1	270083.28	270175.94	273545.63
-1	1	-1	1	-1	-1	-1	1	-1	-1	-1	272570.59	273130.00	275833.47
-1	-1	1	-1	1	-1	-1	1	-1	-1	-1	277644.00	277744.44	275453.75
-1	1	1	-1	-1	1	-1	1	-1	-1	-1	273914.78	273089.50	272481.97
-1	-1	-1	-1	-1	-1	1	1	-1	-1	-1	282453.75	282164.84	282576.19
-1	1	-1	-1	1	1	1	1	-1	-1	-1	272598.69	272831.00	272052.38
-1	-1	1	1	-1	1	1	1	-1	-1	-1	275134.81	276479.47	274012.63
-1	1	1	1	1	-1	1	1	-1	-1	-1	272995.66	272986.84	272015.53
-1	-1	-1	1	-1	1	-1	-1	1	-1	-1	279808.19	279251.34	278463.38
-1	1	-1	1	1	-1	-1	-1	1	-1	-1	276978.41	277362.06	274477.59
-1	-1	1	-1	-1	-1	-1	-1	1	-1	-1	280626.34	281349.03	283072.91
-1	1	1	-1	1	1	-1	-1	1	-1	-1	273965.31	273058.50	272286.69
-1	-1	-1	-1	1	-1	1	-1	1	-1	-1	284488.38	284249.03	281871.84
-1	1	-1	-1	-1	1	1	-1	1	-1	-1	276674.31	276447.94	278073.81
-1	-1	1	1	1	1	1	-1	1	-1	-1	274631.94	274242.44	272161.03
-1	1	1	1	-1	-1	1	-1	1	-1	-1	277075.63	276338.91	276190.97
1	-1	-1	1	-1	-1	-1	1	1	-1	-1	273376.38	273977.06	274504.50
1	1	-1	1	1	1	-1	1	1	-1	-1	272160.13	272903.66	273301.03
1	-1	1	-1	-1	1	-1	1	1	-1	-1	269848.13	268495.13	272287.09
1	1	1	-1	1	-1	-1	1	1	-1	-1	269188.69	267365.28	269757.84
1	-1	-1	-1	1	1	1	1	1	-1	-1	272380.59	272371.56	270562.81
1	1	-1	-1	-1	-1	1	1	1	-1	-1	274011.84	272677.84	273875.00
1	-1	1	1	1	-1	1	1	1	-1	-1	272288.75	273286.66	271861.19
1	1	1	1	-1	1	1	1	1	-1	-1	271703.63	272262.34	273023.16
-1	-1	-1	-1	1	1	-1	-1	-1	1	-1	278668.09	278596.19	277745.78
-1	1	-1	-1	-1	-1	-1	-1	-1	1	-1	281917.41	282506.78	282057.44
-1	-1	1	1	1	-1	-1	-1	-1	1	-1	272411.94	272246.94	275982.03
-1	1	1	1	-1	1	-1	-1	-1	1	-1	271475.09	271670.84	272828.09
-1	-1	-1	1	-1	-1	1	-1	-1	1	-1	282052.63	281011.19	282387.00
-1	1	-1	1	1	1	1	-1	-1	1	-1	271320.00	270516.91	271322.16
-1	-1	1	-1	-1	1	1	-1	-1	1	-1	276208.63	276268.44	279483.06
-1	1	1	-1	1	-1	1	-1	-1	1	-1	277279.38	278598.91	275363.31
1	-1	-1	-1	1	-1	-1	1	-1	1	-1	273663.25	273341.53	273864.41
1	1	-1	-1	-1	1	-1	1	-1	1	-1	273276.91	273578.56	270851.00

TABLE A.1 — CONTINUED

$x^2, x^4, x^7, x^{10}, x^{12}, x^{13}$ $x^{14}, x^{21}, x^{37}, x^{51}, x^{62}$											$\hat{Z}_{RS}(\mathbf{x}_k)$	$\hat{Z}_{CV}(\mathbf{x}_k)$	$\hat{Z}_{LH}(\mathbf{x}_k)$
1	-1	1	1	1	1	-1	1	-1	1	-1	277663.72	281379.84	277887.09
1	1	1	1	-1	-1	-1	1	-1	1	-1	269996.59	271297.16	270683.16
1	-1	-1	1	-1	1	1	1	-1	1	-1	273338.19	272963.75	271529.03
1	1	-1	1	1	-1	1	1	-1	1	-1	270203.00	273442.84	270532.38
1	-1	1	-1	-1	-1	1	1	-1	1	-1	271198.97	270923.31	274892.81
1	1	1	-1	1	1	1	1	-1	1	-1	278212.16	278307.97	275093.44
1	-1	-1	-1	-1	-1	-1	-1	1	1	-1	282820.13	282816.97	281126.63
1	1	-1	-1	1	1	-1	-1	1	1	-1	267157.88	266390.00	269838.88
1	-1	1	1	-1	1	-1	-1	1	1	-1	271664.19	272222.56	271625.06
1	1	1	1	1	-1	-1	-1	1	1	-1	271524.50	271403.19	268896.81
1	-1	-1	1	1	1	1	-1	1	1	-1	273551.25	273914.41	270289.19
1	1	-1	1	-1	-1	1	-1	1	1	-1	275612.81	275430.56	274236.25
1	-1	1	-1	1	-1	1	-1	1	1	-1	270732.84	270923.06	274300.59
1	1	1	-1	-1	1	1	-1	1	1	-1	272924.81	273094.56	271315.56
-1	-1	-1	-1	-1	1	-1	1	1	1	-1	276751.88	275863.72	276010.91
-1	1	-1	-1	1	-1	-1	1	1	1	-1	272564.56	272284.56	272252.25
-1	-1	1	1	-1	-1	-1	1	1	1	-1	274431.34	273284.44	274364.41
-1	1	1	1	1	1	-1	1	1	1	-1	277532.47	277129.44	274888.78
-1	-1	-1	1	1	-1	1	1	1	1	-1	273819.63	274121.69	272788.19
-1	1	-1	1	-1	1	1	1	1	1	-1	267480.81	267363.72	270872.63
-1	-1	1	-1	1	1	1	1	1	1	-1	272635.06	273049.06	270784.69
-1	1	1	-1	-1	-1	1	1	1	1	-1	273861.03	274057.19	273490.38
-1	-1	-1	-1	-1	1	-1	-1	-1	-1	1	280768.50	281276.47	282136.56
-1	1	-1	-1	1	-1	-1	-1	-1	-1	1	276869.47	276409.56	278486.56
-1	-1	1	1	-1	-1	-1	-1	-1	-1	1	280619.06	280726.59	280595.09
-1	1	1	1	1	1	-1	-1	-1	-1	1	271124.44	272042.03	271577.72
-1	-1	-1	1	1	-1	1	-1	-1	-1	1	275369.59	276431.94	278664.19
-1	1	-1	1	-1	1	1	-1	-1	-1	1	273584.34	273295.66	275431.22
-1	-1	1	-1	1	1	1	-1	-1	-1	1	271785.16	271603.53	275384.75
-1	1	1	-1	-1	-1	1	-1	-1	-1	1	279768.41	279507.44	280090.22
1	-1	-1	-1	-1	-1	-1	1	-1	-1	1	277305.41	277003.03	278140.50
1	1	-1	-1	1	1	-1	1	-1	-1	1	276261.97	276185.06	270848.19
1	-1	1	1	-1	1	-1	1	-1	-1	1	276775.63	277764.69	273649.69
1	1	1	1	1	-1	-1	1	-1	-1	1	274530.00	274086.09	273728.31
1	-1	-1	1	1	1	1	1	-1	-1	1	272895.50	272417.06	273177.44
1	1	-1	1	-1	-1	1	1	-1	-1	1	272288.00	271969.41	271461.47
1	-1	1	-1	1	-1	1	1	-1	-1	1	273157.72	272729.75	271950.41
1	1	1	-1	-1	1	1	1	-1	-1	1	270103.81	270765.25	270262.00
1	-1	-1	-1	1	-1	-1	-1	1	-1	1	276202.16	277332.16	277291.84
1	1	-1	-1	-1	1	-1	-1	1	-1	1	275069.81	274744.00	273809.66
1	-1	1	1	1	1	-1	-1	1	-1	1	269949.88	271796.31	271299.13
1	1	1	1	-1	-1	-1	-1	1	-1	1	276457.81	276699.44	272069.50
1	-1	-1	1	-1	1	1	-1	1	-1	1	274575.00	274601.72	274165.84
1	1	-1	1	1	-1	1	-1	1	-1	1	270072.69	271580.91	271040.44
1	-1	1	-1	-1	-1	1	-1	1	-1	1	280422.13	280294.19	278913.28
1	1	1	-1	1	1	1	-1	1	-1	1	268220.09	267690.31	269662.53
-1	-1	-1	-1	1	1	-1	1	1	-1	1	271615.97	271024.00	272807.28

TABLE A.1 — CONTINUED

x^2, x^4, x^7, x^{10}, x^{12}, x^{13} x^{14}, x^{21}, x^{37}, x^{51}, x^{62}											$\hat{Z}_{RS}(\mathbf{x}_k)$	$\hat{Z}_{CV}(\mathbf{x}_k)$	$\hat{Z}_{LH}(\mathbf{x}_k)$
-1	1	-1	-1	-1	-1	-1	1	1	-1	1	275685.03	275554.16	276729.00
-1	-1	1	1	1	-1	-1	1	1	-1	1	268890.28	269699.00	272272.69
-1	1	1	1	-1	1	-1	1	1	-1	1	272479.72	272408.91	272452.13
-1	-1	-1	1	-1	-1	1	1	1	-1	1	274211.63	273738.31	277067.81
-1	1	-1	1	1	1	1	1	1	-1	1	267583.94	267217.88	270455.41
-1	-1	1	-1	-1	1	1	1	1	-1	1	273458.56	272788.25	273689.88
-1	1	1	-1	1	-1	1	1	1	-1	1	271418.94	271980.63	271238.47
1	-1	-1	1	-1	-1	-1	-1	-1	1	1	278552.06	278902.38	278077.56
1	1	-1	1	1	1	-1	-1	-1	1	1	270402.06	270277.66	269766.91
1	-1	1	-1	-1	1	-1	-1	-1	1	1	271874.88	272181.34	275105.56
1	1	1	-1	1	-1	-1	-1	-1	1	1	270349.00	269449.66	271082.63
1	-1	-1	-1	1	1	1	-1	-1	1	1	274688.63	274838.94	272888.22
1	1	-1	-1	-1	-1	1	-1	-1	1	1	278903.19	278913.84	277762.34
1	-1	1	1	1	-1	1	-1	-1	1	1	274149.09	273514.03	272103.28
1	1	1	1	-1	1	1	-1	-1	1	1	268679.03	267704.84	270237.09
-1	-1	-1	1	-1	1	-1	1	-1	1	1	275773.53	276179.81	273628.97
-1	1	-1	1	1	-1	-1	1	-1	1	1	270014.44	270289.34	270501.03
-1	-1	1	-1	-1	-1	-1	1	-1	1	1	277562.13	277240.97	277561.00
-1	1	1	-1	1	1	-1	1	-1	1	1	270610.75	272090.84	270555.66
-1	-1	-1	-1	1	-1	1	1	-1	1	1	281698.16	281097.84	276109.38
-1	1	-1	-1	-1	1	1	1	-1	1	1	274660.88	273686.75	273048.47
-1	-1	1	1	1	1	1	1	-1	1	1	275039.22	278370.06	273290.31
-1	1	1	1	-1	-1	1	1	-1	1	1	268650.34	269092.06	271847.94
-1	-1	-1	1	1	1	-1	-1	1	1	1	272377.22	271524.91	272687.69
-1	1	-1	1	-1	-1	-1	-1	1	1	1	278669.19	279366.28	276824.19
-1	-1	1	-1	1	-1	-1	-1	1	1	1	279659.63	279290.28	277272.72
-1	1	1	-1	-1	1	-1	-1	1	1	1	274125.28	273539.69	273429.53
-1	-1	-1	-1	-1	-1	1	-1	1	1	1	283753.78	283152.31	283831.47
-1	1	-1	-1	1	1	1	-1	1	1	1	272916.91	272017.19	271852.59
-1	-1	1	1	-1	1	1	-1	1	1	1	270736.72	270587.56	274074.19
-1	1	1	1	1	-1	1	-1	1	1	1	271719.44	272061.16	271237.88
1	-1	-1	1	1	-1	-1	1	1	1	1	272845.72	271514.84	269328.56
1	1	-1	1	-1	1	-1	1	1	1	1	268399.06	267801.91	268868.69
1	-1	1	-1	1	1	-1	1	1	1	1	269695.34	270385.41	272674.16
1	1	1	-1	-1	-1	-1	1	1	1	1	268000.31	268926.25	269682.56
1	-1	-1	-1	-1	1	1	1	1	1	1	269832.47	270567.69	272308.13
1	1	-1	-1	1	-1	1	1	1	1	1	269555.75	269303.91	268693.81
1	-1	1	1	-1	-1	1	1	1	1	1	267142.34	267987.06	270404.22
1	1	1	1	1	1	1	1	1	1	1	280877.28	280925.84	280150.44
α	0	0	0	0	0	0	0	0	0	0	262547.16	262891.19	262254.28
$-\alpha$	0	0	0	0	0	0	0	0	0	0	280458.16	281605.63	280449.28
0	α	0	0	0	0	0	0	0	0	0	260381.20	262270.34	257491.28
0	$-\alpha$	0	0	0	0	0	0	0	0	0	268911.56	268994.94	267776.72
0	0	α	0	0	0	0	0	0	0	0	263901.41	266615.88	262171.47
0	0	$-\alpha$	0	0	0	0	0	0	0	0	273256.19	272627.66	271253.25
0	0	0	α	0	0	0	0	0	0	0	266695.69	265269.19	262282.09
0	0	0	$-\alpha$	0	0	0	0	0	0	0	279600.22	279559.66	281936.09

TABLE A.1 — CONTINUED

x^2, x^4, x^7, x^{10}, x^{12}, x^{13} x^{14}, x^{21}, x^{37}, x^{51}, x^{62}											$\hat{Z}_{RS}(\mathbf{x}_k)$	$\hat{Z}_{CV}(\mathbf{x}_k)$	$\hat{Z}_{LH}(\mathbf{x}_k)$
0	0	0	0	α	0	0	0	0	0	0	262767.47	262977.56	266524.84
0	0	0	0	-α	0	0	0	0	0	0	289449.34	289624.91	288697.25
0	0	0	0	0	α	0	0	0	0	0	265740.59	264797.31	267964.00
0	0	0	0	0	-α	0	0	0	0	0	277653.84	278618.81	280118.75
0	0	0	0	0	0	α	0	0	0	0	264566.13	265778.88	261836.02
0	0	0	0	0	0	-α	0	0	0	0	281942.59	281789.44	278625.28
0	0	0	0	0	0	0	α	0	0	0	267259.78	267273.06	265554.47
0	0	0	0	0	0	0	-α	0	0	0	293418.78	293111.38	292761.38
0	0	0	0	0	0	0	0	α	0	0	254724.94	254350.55	254842.89
0	0	0	0	0	0	0	0	-α	0	0	267867.94	268843.72	270128.69
0	0	0	0	0	0	0	0	0	α	0	259339.08	259064.28	256909.66
0	0	0	0	0	0	0	0	0	-α	0	264690.44	264783.06	266351.53
0	0	0	0	0	0	0	0	0	0	α	256759.03	256634.92	257119.61
0	0	0	0	0	0	0	0	0	0	-α	263405.75	264139.34	266553.50
0	0	0	0	0	0	0	0	0	0	0	256393.22	255784.72	256009.81
0	0	0	0	0	0	0	0	0	0	0	255674.83	255718.84	254763.36
0	0	0	0	0	0	0	0	0	0	0	253818.95	253197.25	255341.89
0	0	0	0	0	0	0	0	0	0	0	253555.20	253433.41	254785.86
0	0	0	0	0	0	0	0	0	0	0	252856.48	252810.88	255575.00
0	0	0	0	0	0	0	0	0	0	0	255490.61	255577.50	255235.52
0	0	0	0	0	0	0	0	0	0	0	252902.50	254253.78	255041.83
0	0	0	0	0	0	0	0	0	0	0	253241.92	254443.30	255877.17
0	0	0	0	0	0	0	0	0	0	0	256394.80	256144.66	255931.16
0	0	0	0	0	0	0	0	0	0	0	254077.73	253923.22	255685.38
0	0	0	0	0	0	0	0	0	0	0	254993.14	254751.05	255091.02
0	0	0	0	0	0	0	0	0	0	0	257465.98	257531.95	255311.55
0	0	0	0	0	0	0	0	0	0	0	254823.16	255137.92	255446.63
0	0	0	0	0	0	0	0	0	0	0	255125.48	254839.91	255863.88
0	0	0	0	0	0	0	0	0	0	0	255648.98	255481.08	255543.17
0	0	0	0	0	0	0	0	0	0	0	258436.55	257506.67	254989.86
0	0	0	0	0	0	0	0	0	0	0	257079.67	255383.59	255698.77
0	0	0	0	0	0	0	0	0	0	0	254280.70	254184.72	254915.28
0	0	0	0	0	0	0	0	0	0	0	252949.27	252760.42	255062.73
0	0	0	0	0	0	0	0	0	0	0	260193.55	260242.17	255693.64
0	0	0	0	0	0	0	0	0	0	0	256091.58	256139.03	255246.78
0	0	0	0	0	0	0	0	0	0	0	257106.16	254932.30	255796.36
0	0	0	0	0	0	0	0	0	0	0	254390.89	254333.63	255222.17
0	0	0	0	0	0	0	0	0	0	0	255581.75	255756.41	255611.36
0	0	0	0	0	0	0	0	0	0	0	255913.77	256291.66	255996.63
0	0	0	0	0	0	0	0	0	0	0	257371.08	258470.84	255658.77
0	0	0	0	0	0	0	0	0	0	0	254371.97	254671.95	255388.92

† - '0' codes represent centerpoint values from Table 5.34. Half-ranges are 7.5 for x^2, x^7, x^{10}, x^{14}, and x^{62}; 10 for x^{12}, x^{13}, and x^{21}; and, 5 for x^4, x^{37}, and x^{51}. All $\pm\alpha$ values are ± 3.3555.

Bibliography

Avramidis, Athanassios N. "Variance Reduction for Quantile Estimation via Correlation Induction." In *Proceedings of the 1992 Winter Simulation Conference*, Swain, J. J., R. C. Crain, and J. R. Wilson, eds. New York: Association for Computing Machinery, 1992.

Avramidis, Athanassios N. and James R. Wilson. "Integrated Variance Reduction Strategies for Simulation." North Carolina State University, February 7, 1995. To appear in *Operations Research.*

Avriel, M. and A. C. Williams. "The Value of Information and Stochastic Programming." *Operations Research* 18, no. 5 (September-October 1970): 947-954,

Azadivar, F. "A Tutorial on Simulation Optimization." In *Proceedings of the 1992 Winter Simulation Conference*, Swain, J.J., D. Goldsman, R.C. Crain, and J. R. Wilson, eds. New York: Association of Computing Machinery, 1992.

Baron, David P. "Information in Two-Stage Programming Under Uncertainty." *Naval Research Logistics Quarterly* 18, no. 2 (June 1971): 169-176.

Barton, Russell R. and John S. Ivey. "Modifications of the Nelder-Mead Simplex Method for Stochastic Simulation Response Optimization." In *Proceedings of the 1991 Winter Simulation Conference*, Nelson, Barry W., David Kelton, and Gordon M. Clark, eds. New York: Association for Computing Machinery, 1991.

Bazaraa, Mokhtar S., John J. Jarvis, and Hanif D. Sherali. *Linear Programming and Network Flows.* 2d ed. New York: John Wiley, 1990.

Beale, E. M. L. "On Minimizing a Convex Function Subject to Linear Inequalities." *Journal of the Royal Statistical Society (Series B: Methodology)* 17, no. 2 (1955): 173-184.

Benders, J. F. "Partitioning Procedures for Solving Mixed-Variables Programming Problems." *Numerische Mathematik* 4, no. 2 (May 1962): 238-252.

Bereanu, Bernard. "Some Numerical Methods in Stochastic Linear Programming Under Risk and Uncertainty." In *Stochastic Programming*, ed. M. A. H. Dempster, 169-205. London: Academic Press, 1980.

Berger, Adam J. and John M. Mulvey. "Restarting Strategies for the DQA Algorithm." Princeton, N. J.: Princeton University, Department of Civil Engineering and Applied Science, Technical Report SOR-93-6, April 1993.

Berger, Adam J., John M. Mulvey, and Andrzej Ruszcznski. "A Distributed Scenario Decomposition Algorithm for Large Stochastic Programs." Princeton, N. J.: Princeton University, Department of Civil Engineering and Applied Science, Technical Report SOR-93-2, February 1993. Photocopied.

Biles, William E. and James J. Swain. "Mathematical Programming and the Optimization of Computer Simulations." *Mathematical Programming Study 11*, (1979): 189-207.

Birge, John R. "The Value of the Stochastic Solution in Stochastic Linear Programs with Fixed Recourse." *Mathematical Programming* 23, no. 3 (1982): 314-325.

_____. "Aggregation Bounds in Stochastic Linear Programming." *Mathematical Programming* 31, no. 1 (January 1985): 25-41.

_____. "The Relationship Between the L-Shaped Method and Dual Basis Factorization for Stochastic Linear Programming." In *Numerical Techniques for Stochastic Optimization*, eds. Yuri Ermoliev and Roger J-B Wets, 267-272. Springer Series in Computational Mathematics X. Berlin: Springer-Verlag, 1988.

Birge, John R. and José H. Dulá. "Bounding Separable Recourse Functions with Limited Distribution Information." *Annals of Operations Research* 30, (1991): 277-298.

Birge, John R. and Francois V. Louveaux. "A multicut algorithm for two-stage stochastic linear programs." *European Journal of Operations Research* 34, no. 3 (March 1988): 384-392.

Birge, John R. and John M. Mulvey. "Stochastic Programming in Industrial Engineering, 1994" TMs. Department of Civil Engineering and Operations Research, Princeton University, Princeton, N. J.

Birge, John R. and Stein W. Wallace. "A Separable Piecewise Linear Upper Bound for Stochastic Linear Programs." *SIAM Journal of Control and Optimization* 26, no. 3 (May 1988): 725-739.

Birge, John R. and R. J-B. Wets. "Designing Approximation Schemes for Stochastic Optimization Problems, in Particular for Stochastic programs with Recourse." *Mathematical Programming Study 27*, (1986): 54-102.

_____. "Sublinear Upper Bounds for Stochastic Programs with Recourse." *Mathematical Programming* 43, no. 2 (February 1989): 131-149.

Box, George E. P., and Norman R. Draper. *Empirical Model-Building and Response Surfaces.* New York: John Wiley, 1987.

Box, G. E. P., and J. S. Hunter. "The 2^{k-p} Fractional Factorial Designs Part II." *Technometrics* 3, no. 4 (November 1961): 449-458.

Box, George E. P., William G. Hunter, and J. Stuart Hunter. *Statistics for Experimenters.* New York: John Wiley, 1978.

Bradley, James V. *Distribution-Free Statistical Tests.* Englewood Cliffs, New Jersey: Prentice-Hall, 1968.

Charnes, A. and W. W. Cooper. "Chance-Constrained Programming." *Management Science* 6, no. 1 (October 1959): 73-79.

Conover, W. J. *Practical Nonparametric Statistics.* 2d ed. New York: John Wiley, 1980.

Cornell, John A. "Experiments with Mixtures: A Review." *Technometrics* 15, no. 3 (August 1973): 437-455.

_____. "Experiments with Mixtures: An Update and Bibliography." *Technometrics* 21, no. 1 (February 1979): 95-106.

_____. *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data.* New York: John Wiley, 1981.

Dantzig, George B. "Linear Programming Under Uncertainty." *Management Science* 1, nos. 3 and 4 (April-July 1955): 197-206.

Dantzig, George B. and Peter W. Glynn. "Parallel Processors for Planning Under Uncertainty." *Annals of Operations Research* 22, (1990): 1-21.

Dantzig, George B. and Gerd Infanger. "Large-Scale Stochastic Linear Programs: Importance Sampling and Benders Decomposition." Stanford, California: Stanford University, Systems Optimization laboratory Technical Report SOL 91-4, March, 1991.

Dantzig, George B. and Albert Madansky. "On the Solution of Two-Stage Linear Programs Under Uncertainty." In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability Held at the Statistical Laboratory, University of California, June 20 - July 30, 1960,* edited by Jerzy Neyman, 165-176. Berkeley: University of California Press, 1961.

Dempster, M. A. H. "Introduction to Stochastic Programming." In *Stochastic Programming*, ed. M. A. H. Dempster, 3-59. London: Academic Press, 1980.

Dembo, Ron S. "Scenario Optimization." Toronto: Algorithmics Inc. Research Paper 89.01, July 1989.

_____. "Scenario Optimization." *Annals of Operations Research* 30, (1991): 63-80.

Diamond, William J. *Practical Experimental Designs*. 2d ed. New York: Van Nostrand Reinhold, 1989.

Draper, N. R. and W. E. Lawrence. "Mixture Designs for Three Factors." *Journal of the Royal Statistical Society (Series B: Methodology)* 27, no. 3 (1965): 450-465.

_____. "Mixture Designs for Four Factors." *Journal of the Royal Statistical Society (Series B: Methodology)* 27, no. 3 (1965): 473-478.

Draper, Norman. R. and Dennis K. Lin. "Connections Between Two-Level Designs of Resolution III and V." *Technometrics* 32, no. 4 (August, 1990): 283-288.

Edirisinghe, N. C. P. and W. T. Ziemba. "Tight Bounds for Stochastic Convex Programs." *Operations Research* 40, no. 4 (July-August 1992): 660-677.

Ermoliev, Yuri. "Stochastic Quasigradient Methods and their Application to System Optimization." *Stochastics* 9, no. 1 and 2 (March 1983): 1-36.

_____. "Stochastic Gradient Methods." In *Numerical Techniques for Stochastic Optimization*, eds. Yuri Ermoliev and Roger J-B Wets, 141-185. Springer Series in Computational Mathematics X. Berlin: Springer-Verlag, 1988.

Ermoliev, Yuri. and E. A. Nurminsky. "Stochastic Quasigradient Algorithms for Minimax Problems in Stochastic Programming." In *Stochastic Programming*, ed. M. A. H. Dempster, 275-285. London: Academic Press, 1980.

Ermoliev, Yuri and Roger J-B Wets. "Stochastic Programming, an Introduction." In *Numerical Techniques for Stochastic Optimization*, eds. Yuri Ermoliev and Roger J-B Wets, 1-32. Springer Series in Computational Mathematics X. Berlin: Springer-Verlag, 1988.

Frauendorfer, Karl. "Solving SLP Recourse Problems with Arbitrary Multivariate Distributions -- The Dependent Case." *Mathematics of Operations Research* 13, no. 3 (August 1988): 377-394.

Fu, Michael C. "Optimization via Simulation: A review." *Annals of Operations Research* 53, (1994): 199-247.

Gaivoronski, A. "Stochastic Quasigradient Methods and Their Implementation." In *Numerical Techniques for Stochastic Optimization*, eds. Yuri Ermoliev and Roger J-B Wets, 313-351. Springer Series in Computational Mathematics X. Berlin: Springer-Verlag, 1988.

Gartska, Stanley J. "The Economic Equivalence of Several Stochastic Programming Models." In *Stochastic Programming*, ed. M. A. H. Dempster, 83-91. London: Academic Press, 1980.

Gartska, Stanley J. and David P. Rutenberg. "Computation in Discrete Stochastic Programs with Recourse." *Operations Research* 21, no. 1 (January-February 1973): 112-122.

Gassmann, H. and W. T. Ziemba. "A Tight Upper Bound for the Expectation of a Convex Function of a Multivariate Random Variable." *Mathematical Programming Study 27*, (1986): 39-53.

Gill, P.E., W. Murray, and M. H. Wright. *Practical Optimization*. London: Academic Press, 1981. Quoted in David G. Luenberger. *Linear and Nonlinear Programming*. 2d ed. Reading, Massachusetts: Addison-Wesley, 1989.

Glynn, Peter. E-mail correspondence to Thomas G. Bailey, University of Texas, Austin, October 6, 1994.

Hansotia, Behram J. "Stochastic Linear Programming with Recourse: A Tutorial." *Decision Sciences* 11, no. 1 (January 1980): 151-168.

Hartley, H. O. "Smallest Composite Designs for Quadratic Response Surfaces." *Biometrics* 15, no. 4 (December 1959): 611-624

Haugland, Dag and Stein W. Wallace. "Solving Many Linear Programs that Differ Only in the Righthand Side." *European Journal of Operational Research* 37, no. 3 (December 1988): 318-324.

Hausch, D. B. and W. T. Ziemba. "Bounds on the Value of Information in Uncertain Decision Problems II." *Stochastics* 10, (1983): 181-217.

Helgason, Thorkell and Stein W. Wallace. "Approximate Scenario Solutions in the Progressive Hedging Algorithm." *Annals of Operations Research* 31, (1991): 425-444

Higle, Julia L. and Suvrajeet Sen. "Finite Master Programs in Regularized Stochastic Decomposition." Department of Systems and Industrial Engineering, University of Arizona, 1990. Quoted from Holmes, Derek. *Stochastic Programming Problem Library*. WWW page at http://www.engin.umich.edu/~dholmes/SPTSlists.html, Department of Industrial Engineering and Operations Research, University of Michigan, 9 May 1995.

_____. "Statistical Verification of Optimality Conditions for Stochastic Programs with Recourse." *Annals of Operations Research* 30, (1991): 215-240.

_____. "Stochastic Decomposition: An Algorithm for Two-Stage Linear Programs with Recourse." *Mathematics of Operations Research* 16, no. 3 (August 1991): 650-679

_____. "Duality and Statistical Tests of Optimality for Two Stage Stochastic Programs." Tucson, Arizona: University of Arizona, SIE Department Working Paper, August, 1993.

Hoerl, A. E. "Optimum Solution of Many Variables Equations." *Chem. Eng. Prog.* 55, (1959): 69-78. Quoted in Box, George E. P., and Norman R. Draper. *Empirical Model-Building and Response Surfaces.* New York: John Wiley, 1987.

Hoerl, R. W. "Ridge Analysis 25 Years Later." *Am. Statist.* 39, (1985): 186-192. Quoted in Box, George E. P., and Norman R. Draper. *Empirical Model-Building and Response Surfaces.* New York: John Wiley, 1987.

Holmes, Derek. *Stochastic Programming Problem Library*. WWW page at http://www.engin.umich.edu/~dholmes/SPTSlists.html, Department of Industrial Engineering and Operations Research, University of Michigan, 9 May 1995.

Infanger, Gerd. *Planning Under Uncertainty*. Danvers, Mass: Boyd and Fraser, 1994.

International Business Machines Corp. *Optimization and Tuning Guide for the XL FORTRAN and XL C Compilers (AIX Version 3.2 for RISC System 6000).* North York, Ontario: International Business Machines Corp., September, 1992.

Jacobson, Sheldon H. and Lee W. Schruben. "Techniques for Simulation Response Optimization." *Operations Research Letters* 8, no. 1 (February 1989): 1-9.

Kall, Peter. *Stochastic Linear Programming*. Economics and Operations Research XXI. Berlin: Springer-Verlag, 1976.

_____. "Computational Methods for Solving Two-Stage Stochastic Linear Programs." *Zeitschrift für angewandte Mathematik und Physik* 30, no. 2 (June 1979): 261-270.

Kall, P., A. Ruszczynski, and K. Frauendorfer. "Approximation Techniques in Stochastic Programming." In *Numerical Techniques for Stochastic Optimization*, eds. Yuri Ermoliev and Roger J-B Wets, 33-64. Springer Series in Computational Mathematics X. Berlin: Springer-Verlag, 1988.

Kall, Peter, and Stein W. Wallace. *Stochastic Programming*. New York: John Wiley, 1994.

Kelley, J. E. "The Cutting-Plane Method for Solving Complex Programs." *Journal of the SIAM* 8, no. 4 (December 1960): 703-712.

Kirby, M. J. L. "The Current State of Chance-Constrained Programming." In *Proceedings of the Princeton Symposium on Mathematical Programming*, edited by Harold W. Kuhn, 93-111. Princeton, N.J.: Princeton University, 1970.

Kleijnen, Jack P. C. *Statistical Techniques in Simulation Part I.* Statistics: Textbooks and Monographs Vol. 9. New York: Marcel Dekker, 1974.

_____. *Statistical Tools for Simulation Practitioners.* Statistics: Textbooks and Monographs Vol. 76. New York: Marcel Dekker, 1987.

Lavenberg, Stephen S., Thomas L. Moeller, and Peter D. Welch. "Statistical Results on Control Variables with Application to Queueing Network Simulation." *Operations Research* 30, no. 1 (January-February 1982): 182-202.

Lavenberg, Stephen S. and Peter Welch. "A Perspective on the Use of Control Variables to Increase the Efficiency of Monte Carlo Simulations." *Management Science* 27, no. 3 (March 1981): 322-335.

Law, Averill M. and W. David Kelton. *Simulation Modeling and Analysis*. 2d ed. New York: McGraw-Hill, 1991.

Lorenzen, Thomas J. and Virgil L. Anderson. *Design of Experiments: A No-Name Approach.* Statistics: Textbooks and Monographs Vol. 139. New York: Marcel Dekker, 1993.

Luenberger, David G. *Linear and Nonlinear Programming.* 2d ed. Reading, Massachusetts: Addison-Wesley, 1989.

Madansky, Albert. "Inequalities for Stochastic Linear Programming Problems." *Management Science* 6, no. 2 (January 1960): 197-204.

Marti, K. "Descent Stochastic Quasigradient Methods." In *Numerical Techniques for Stochastic Optimization*, eds. Yuri Ermoliev and Roger J-B Wets, 393-401. Springer Series in Computational Mathematics X. Berlin: Springer-Verlag, 1988.

McKay, Michael D. "Sensitivity and Uncertainty Analysis Using a Statistical Sample of Input Values." In *Uncertainty Analysis*, ed. Yigal Ronen, 146-185. Boca Raton, Florida: CRC Press, 1988.

McKay, Michael D. "Latin Hypercube Sampling as a Tool in Uncertainty Analysis of Computer Models." *Proceedings of the 1992 Winter Simulation Conference*, Swain, J. J., D. Goldsman, R. C. Crain, and J. R. Wilson, eds. New York: Association for Computing Machinery, 1992.

McKay, M. D., W. J. Conover and R. J. Beckman. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code." *Technometrics* 21, no. 2, (May 1979): 239-245.

Meketon, Marc S. "Optimization in Simulation: A Survey of Recent Results." In *Proceedings of the 1987 Winter Simulation Conference*, Thesen, Arne, Hank Grant, and W. David Kelton, eds. New York: Association for Computing Machinery, 1987.

Miller, Rupert G. "The Jackknife - A Review." *Biometrika* 61, no. 1, (April 1974): 1-15.

Montgomery, Douglas C. *Design and Analysis of Experiments.* 2d ed. New York: John Wiley, 1984.

Morris, J. G. and H. E. Thompson. "A Note on the 'Value' of Bounds on EVPI in Stochastic Linear Programming." *Naval Research Logistics Quarterly* 27, no. 1 (March 1980): 165-169.

Morris, M. D. "Factorial Sampling Plans for Preliminary Computational Experiments." *Technometrics* 33, no. 2, (May, 1991): 161-174.

Morton, David P. "Stopping Rules for a Class of Sampling-Based Stochastic Programming Algorithms." Monterey, California: Naval Postgraduate School Report NPS-OR-94-003, 1994.

_____. "Stopping Rules for a Class of Sampling-Based Stochastic Programming Algorithms," Seminar presented to the Department of Mechanical Engineering Operations Research Group, University of Texas, Austin, November 11, 1994.

_____. Problem set provided by e-mail correspondence to Thomas G. Bailey, University of Texas, Austin, November 25, 1994.

_____. Conversation with author. University of Texas, Austin, February 20, 1995.

_____. "A Sampling-Based Algorithm for Two-Stage Stochastic Linear Programming." Working paper, Department of Mechanical Engineering, University of Texas, Austin, September, 1995.

Mulvey, John M., Robert J. Vanderbei, and Stavros A. Zenios. "Robust Optimization of Large Scale Systems." Princeton, N. J.: Princeton University, Statistics and Operations Research Report SOR-91-13, 1991. Photocopied.

Mulvey, John M. and Hercules Vladimirou. "Applying the Progressive Hedging Algorithm to Stochastic Generalized Algorithms." *Annals of Operations Research 31*, (1991): 399-424.

_____. "Stochastic Network Programming for Financial Planning Problems." *Management Science* 38, no. 11 (November 1992): 1642-1663.

Murty, Katta G. *Linear Programming*. New York: John Wiley, 1983.

Myers, Jerome L. *Fundamentals of Experimental Design*. 3d ed. Boston: Allyn and Bacon, 1979.

Nelder, J. A. and R. Mead. "A Simplex Method for Function Minimization." *The Computer Journal* 7, no. 4 (January 1965): 308-311.

Nelson, Barry L. "On Control Variate Estimators." *Computers in Operations Research* 14, no. 3 (March 1987): 219-225.

_____. "Variance Reduction for Simulation Practitioners." *Proc. 1987 Winter Simulation Conference*, Atlanta, (1987): 43-51. Quoted in Law, Averill M., and W. David Kelton. *Simulation Modeling and Analysis*. 2d ed. New York: McGraw-Hill, 1991.

_____. "Control Variate Remedies." *Operations Research* 38, no. 6 (November-December 1990)

Patel, Nitin R. "On Wait-and-See Stochastic Linear Programmes: An Application and an Algorithm." *Journal of the Operational Research Society* 31, no. 8 (August 1980): 733-741.

Pflug, G. Ch. "Stepsize Rules, Stopping Times and their Implementation in Stochastic Quasigradient Algorithms." In *Numerical Techniques for Stochastic Optimization*, eds. Yuri Ermoliev and Roger J-B Wets, 353-372. Springer Series in Computational Mathematics X. Berlin: Springer-Verlag, 1988.

Plackett, R. L. and J. P. Burman. "The Design of Optimum Multifactorial Experiments." *Biometrika* 33, (1946): 305-325.

Prekopa, A. "On Probabilistic Constrained Programming." In *Proceedings of the Princeton Symposium on Mathematical Programming*, edited by Harold W. Kuhn, 113-138. Princeton, N.J.: Princeton University, 1970.

Pritsker, A. Alan B. *Introduction to Simulation and SLAM*. 3d ed. New York: John Wiley, 1986.

Rockafeller, R. T. and R. J-B. Wets. "A Note About Projections in the Implementation of Stochastic Quasigradient Methods." In *Numerical Techniques for Stochastic Optimization*, eds. Yuri Ermoliev and Roger J-B Wets, 385-392. Springer Series in Computational Mathematics X. Berlin: Springer-Verlag, 1988.

_____. "Scenarios and Policy Aggregation in Optimization Under Uncertainty." *Mathematics of Operations Research* 16, no. 1 (February 1991): 119-147.

Rosen, J. B. "The Gradient Projection Method for Nonlinear Programming. Part I. Linear Constraints." *J. Soc. Indust. Math.* 8, no. 1 (March 1960): 181-217.

Rubenstein, Reuven Y. *Simulation and the Monte Carlo Method*. New York: John Wiley, 1981.

Rubenstein, Reuven Y. and Ruth Marcus. "Efficiency of Multivariate Control Variates in Monte Carlo Simulation." *Operations Research* 33, no. 3 (May-June 1985): 661-676.

Ruszczynski, Andrzej and Wojciech Syski. "A Method of Aggregate Stochastic Subgradients with On-Line Stepsize Rules for Convex Stochastic Programming Problems." *Mathematical Programming Study 28*, (1986): 113-131.

SAS Institute. *User's Guide, Vol. 2, GLM-VARCOMP.* Ver. 6, 4th ed. Cary, North Carolina: SAS Institute, Inc., 1990.

Safizadeh, M. Hossein. "Optimization in Simulation: Current Issues and the Future Outlook." *Naval Research Logistics* 37, no. 6 (December 1990): 807-825.

Scheffe, H. "Experiments with Mixtures." *Journal of the Royal Statistical Society (Series B: Methodology)* 20, no. 2 (1958): 344-360.

Sen, Suvrajeet, Jason Mai, and Julia Higle. "Solution of Large Scale Stochastic Programs with Stochastic Decomposition Algorithms." In *Large Scale Optimization: State of the Art,* eds. W. W. Hager et al., 388-410. Amsterdam: Kluwer, 1994.

Spendley, W., G. R. Hext, and F. R. Himsworth. "Sequential Applications of Simplex Designs in Optimisation and Evolutionary Operation." *Technometrics* 4, no. 4 (November 1962): 441-461.

Sivazlian, B. D. and L. E. Stanfel. *Optimization Techniques in Operations Research.* Englewood Cliffs, New Jersey: Prentice-Hall, 1975.

Stancu-Minasian, I.M., and M. J. Wets. "A Research Bibliography in Stochastic Programming, 1955-1975." *Operations Research* 24, no. 6 (November-December 1976): 1078-1119.

Stein, Michael. "Large Sample Properties of Simulations Using Latin Hypercube Sampling." *Technometrics* 29, no. 2, (May 1987): 143-151.

Steinberg, David M. and William G. Hunter. "Experimental Design: Review and Comment". *Technometrics* 28, no. 2, (May 1984): 71-97.

Strazicky, Beáta. ""On an Algorithm for Solution of the Two-Stage Stochastic Programming Problem." *Methods of Operations Research XIX*, (1974): 142-156.

_____. "Some Results Concerning an Algorithm for the Discrete Recourse Problem." In *Stochastic Programming*, ed. M. A. H. Dempster, 263-274. London: Academic Press, 1980.

Thompson, W. O. and R. H. Myers. "Response Surface Designs for Experiments with Mixtures." *Technometrics* 10, no. 4 (1968): 739-756.

Tocher, K. D. *The Art of Simulation*. London: The English Universities Press, 1963. Quoted in Kleijnen, Jack P. C. *Statistical Techniques in Simulation Part I.* Statistics: Textbooks and Monographs Vol. 9. New York: Marcel Dekker, 1974.

Urasiev, S. "Adaptive Stochastic Quasigradient Procedures." In *Numerical Techniques for Stochastic Optimization*, eds. Yuri Ermoliev and Roger J-B Wets, 373-384. Springer Series in Computational Mathematics X. Berlin: Springer-Verlag, 1988.

Vajda, S. *Probabilistic Programming*. New York: Academic Press, 1972.

Van Slyke, R. M., and Roger Wets. "*L*-Shaped Linear Programs with Applications to Optimal Control and Stochastic Programming." *SIAM Journal of Applied Mathematics* 17, no. 4 (July 1969): 638-663.

Walkup, David M., and Roger J.-B. Wets. "Stochastic Programs with Recourse." *SIAM Journal of Applied Mathematics* 15, no. 5 (September 1967): 1299-1314.

Wallace, Stein W. "Decomposing the Requirement Space of a Transportation Problem into Polyhedral Cones." *Mathematical Programming Study* 28 (1986): 29-47.

Wets, Roger J.-B. "Programming Under Uncertainty: The Equivalent Convex Program." *SIAM Journal of Applied Mathematics* 14, no. 1 (January 1966): 89-105.

_____. "Programming Under Uncertainty: The Solution Set." *SIAM Journal of Applied Mathematics* 14, no. 5 (September 1966): 1143-1151.

_____. "Characterization Theorems for Stochastic Programs." *Mathematical Programming* 2, no. 2 (April 1972): 166-175

_____. "Stochastic Programs with Fixed Recourse: The Equivalent Deterministic Program." *SIAM Review* 16, no. 3 (July 1974): 309-339.

_____. "Stochastic Programming: Solution Techniques and Approximation Schemes." In *Mathematical Programming: The State of the Art,* eds. A. Bachem, M. Grotschel, and B. Korte, 566-603. Berlin: Springer-Verlag, 1983.

_____. "Large Scale Linear Programming Techniques." In *Numerical Techniques for Stochastic Optimization*, eds. Yuri Ermoliev and Roger J-B Wets, 65-93. Springer Series in Computational Mathematics X. Berlin: Springer-Verlag, 1988.

_____. "The Aggregation Principle in Scenario Analysis and Stochastic Optimization." In *Algorithms and Model Formulations*, ed. Stein W. Wallace, 91-113. NATO ASI Series F: Computer and Systems Science 51. Berlin: Springer-Verlag, 1989.

Whitwell, John C., and Graham K. Morbey. "Reduced Designs of Resolution V." *Technometrics* 3, no. 4 (November 1961): 459-477.

Wilson, James R. "Variance Reduction Techniques for Digital Simulation." *Am. J. Math. Management Sci. 4* (1984): 277-312. Quoted in Law, Averill M., and W. David Kelton. *Simulation Modeling and Analysis*. 2d ed. New York: McGraw-Hill, 1991.

_____. Conversation with author. University of Texas, Austin, April 28, 1995.

Vita

Thomas Glenn Bailey was born February 23, 1956 in Lubbock, Texas, the son of Tommy Ross Bailey and Reba Jean Bailey. After graduating from Permian High School, Odessa, Texas, in 1974 he entered the United States Air Force Academy, where he received his commission and Bachelor of Science in May, 1978. After completing Undergraduate Pilot Training in 1979 at Vance AFB, Oklahoma, he remained there as a T-37 instructor pilot and T-37 / T-38 maintenance check pilot while earning a Master of Arts in Political Science from Oklahoma State University in 1982. In 1983, he was reassigned to Travis AFB, California, where he served as a C-5 transport pilot and wing staff officer until entering the Air Force Institute of Technology resident program in 1987. After completing his Master of Science in Operations Research in 1988, he was assigned to the Air Force Military Personnel Center, Randolph AFB, Texas, as a Rated Force Analyst and Chief of the Force Analysis Branch. He entered the University of Texas at Austin in August, 1992. In 1983 he married Beverly K. Massey of Odessa, Texas. They have a daughter Rebecca and a son Kenneth.

Permanent address: 1305 Cordova, Odessa, Texas 79762

This dissertation was typed by the author.